

QY 298 FFYKGINITVTVYTGVLKSGAKPGSADLPENTNIATINPNTSNDPQGVKTVVRDGOITIK 357  
DB 429 -----TLKFYVYFHLNEKADPTKG-----FKNEANVDNGHTDDQTPPTVEVVTGKGRFI 477  
QY 358 KIDGSTKA--SLOGAIFVLKNA--TGQFLNFDNTNN--VEW--GTEANATEYTTGADGIIT 410  
DB 478 KVDGVTATQALAGASFVVRDQNSDTANYLKIDETTKAATWVKTKABEATTTTADGLVD 537  
QY 411 ITGLKEGYTYLVEKKAPGLNLDNSQKVLGDGATDTTNSDNLVNPVTVEN--NKGTELP 469  
DB 538 ITGLKYGTYYLEETVAPDDVLLTNRIEFVNEQSYGTT--ENLVSPEKVPNKHGKT-LP 594  
QY 470 STGGIGTTTFYIIGAILVIGAGIVLVARR 499  
DB 595 STGGRGIYVYLSGAVLLLIAGVYFARRK 624  
QY 470 STGGIGTTTFYIIGAILVIGAGIVLVARR 499  
DB 595 STGGRGIYVYLSGAVLLLIAGVYFARRK 624

## RESULT 3

US-10-661-809-13  
; Sequence 13, Application US/10661809  
; Publication No. US2004010191A1  
; GENERAL INFORMATION:  
; APPLICANT: HOOK, Magnus  
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR  
; FILE REFERENCE: P07741US01/BAS  
; CURRENT FILING DATE: 2003-09-15  
; PRIOR APPLICATION NUMBER: 60/410303  
; PRIOR FILING DATE: 2002-09-13  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 627  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-661-809-13

Query Match 13.0%; Score 334; DB 16; Length 627;  
Best Local Similarity 26.3%; Pred. No. 5e-18;  
Matches 150; Conservative 79; Mismatches 223; Indels 118; Gaps 26;  
QY 3 KMIQSLVLASLAFGMVSPVTPPIAFAA--ETGTTVQDTQKATYKAYKVPDAIDNAN 60  
DB 100 KOAVQSL-----TPGKPVAGQTTDANGNVTVLPKQKQKDAVYTIKEBPKEGV 148  
QY 61 VSDSNKOGA--SYLIPQGEAEYKASTDFNSLF-----TTTNGGRTVYTKDTASANE- 112  
DB 149 VAATNNVAPPVYEMIKQDGSYKYGTEELAVVHIYPRKNVANDGSLHVKVGTAE--NEG 207  
QY 113 -----IATWA-----KSIANTTPV--STVTESNND 136  
DB 208 LNAEFVSKSGSPCTVKYIOGVKDLTYTTDKEQAKRFITGKSYELGENDFTEAEN- 266  
QY 137 GTEVINVS--QGYVYVS--STVNGAVIMVTSVTP-----NATHEK--NTDWTGSG 184  
DB 267 GTGELTVKNLEGVSYLIEVKAPNNAELTENQTKPTFTTEANNQTPVEKTVKNDTSKVDK 326  
QY 185 GCKTVQDKTYSVGDVTKYITIKNAVNYHGTE-----KYQYVVKDTPMSASVVD--LNEGS 239  
DB 327 TTPSLDGKDAIAGEIKYQISVNIPLIGIADKGDANKYVKNLVDKDAALTFDNTSGE 386  
QY 240 YEVTTIDSGNITLTQGESEKATGYNLLENNNFITII--PWAATNTPGTNGTONGANDD 297  
DB 387 YAYALVDGTVI-----APENYQVTEQANGFTAVNPAYIPTLPFG----- 428  
QY 298 FFYKGINITVTVYTGVLKSGAKPGSADLPENTNIATINPNTSNDPQGVKTVVRDGOITIK 357  
DB 429 -----TLKFYVYFHLNEKADPTKG-----FKNEANVDNGHTDDQTPPTVEVVTGKGRFI 477  
QY 358 KIDGSTKA--SLOGAIFVLKNA--TGQFLNFDNTNN--VEW--GTEANATEYTTGADGIIT 410

DB 478 KVDGVTATQALAGASFVVRDQNSDTANYLKIDETTKAATWVKTKABEATTTTADGLVD 537  
QY 411 ITGLKEGYTYLVEKKAPGLNLDNSQKVLGDGATDTTNSDNLVNPVTVEN--NKGTELP 469  
DB 538 ITGLKYGTYYLEETVAPDDVLLTNRIEFVNEQSYGTT--ENLVSPEKVPNKHGKT-LP 594  
QY 470 STGGIGTTTFYIIGAILVIGAGIVLVARR 499  
DB 595 STGGRGIYVYLSGAVLLLIAGVYFARRK 624

## RESULT 4

US-10-661-809-19  
; Sequence 19, Application US/10661809  
; Publication No. US2004010191A1  
; GENERAL INFORMATION:  
; APPLICANT: HOOK, Magnus  
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR  
; FILE REFERENCE: P07741US01/BAS  
; CURRENT FILING DATE: 2003-09-15  
; PRIOR APPLICATION NUMBER: 60/410303  
; PRIOR FILING DATE: 2002-09-13  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 625  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-661-809-19

Query Match 12.3%; Score 318; DB 16; Length 625;  
Best Local Similarity 24.2%; Pred. No. 9.7e-17;  
Matches 141; Conservative 91; Mismatches 206; Indels 144; Gaps 28;  
QY 3 KMIQSLVLASLAFGMVSPVTPPIAFAA--ETGTTV-----QDTQKATYKAYKVPDAI 56  
DB 98 KOAVQSL-----TPGTPVASGTTDADGNVTVLSLPKKQKQKDAVY----- 136  
QY 57 DNANVSDSNKDGAS-----YLIPOGKEAEYKAST---DFNSLF--TTTNGGRT 100  
DB 137 ---TIKEBPKQGVSAANVLAFPVYEMIKQADGSYKYGTEELDTIHLVPRKNTVNDGTL 193  
QY 101 YVTKDTASANEIATWAKSISANT--TP-----SVTVTE----- 132  
DB 194 KVTIKIGTAE--NEALNGAEFLISKEEGTSPVKYIQSVTDGLYTTTDDTKAKHFITGHSY 252  
QY 133 ---SNND-----GTEVINVSQGYVYVS--STVNGAVIMVTSVTPNATHEKVT--- 177  
DB 253 DIGNDFAEASIEKGQLIVNHLEVKYNLEEVKAPDNAAEMIEKQITPPEILANSQTPVE 312  
QY 178 -----DATWGGGGKTVQDKTYSVGDVTKYITIKYK---NAVNYHGTEKYV-QYVVKDTP 228  
DB 313 KTIKNDTSKVDKTPQLNGKQDAIAGEIKYQISEVNIPLIGIADKGTQNKYTTTFKUIDTHD 372  
QY 229 SASVVVDL--NEGSYEVTIITDGSNITLTQGESEKATGYNLLENNNFITIIIP--WAAATNT 285  
DB 373 AALTFDNDSSGYAYALVDGKNEIDPV-----NYSVTEQTDGFTVSDVPNPISLT 423  
QY 286 PTGNTQNGANDDFYKGINITVTVYTGVLKSGAKPGSADLPENTNIATINPNTSNDPQGV 345  
DB 424 PGG-----TLKFVYVYHMLNEKADPTKG-----FSNQANVDNGHTNDQTPP 463  
QY 346 KTVVRDGOITIKIDG--STRASLOGAIFVLKNA---TGQFLNFDNTNNVEM--GTEANA 398  
DB 464 SVDVVTGGRFVKVDGVDVTSQTLAGAEFVVRDQSDTAKYLSIDPSTKAVSWSAKESA 523  
QY 399 TEYTTGADGIITITGLKEGYTYLVEKKAPGLNLDNSQKVLGDGATDTTNSDNLVNP 458  
DB 524 TVFTTNSGLIDVTLGLKYGTYYLEETKAPEKVPVLTNRVAFVITIDEQSYVTAQ---LISP 580  
QY 459 --TVENNKGTLPSTGGIGTTTFYIIGAILVIGAGIVLVARR 498





Db 355 TCAPIPAGAEATFDLVAQTG-----KVQTVTLTTDKNTVTVNGLDKNTYKFEVRS 407  
QY 318 AKPGSADLPENTNIATI-----NPNTSNDPQKQVTVRDGQITIKKIDGSKASLOQAI 371  
Db 408 IKGSADYQEIITAGELAVKWKNDENPKPLDPTEPKVTVYGGKFKVNDKDNK--LAGAE 465  
QY 372 FVLKNA--TGQFL----- 382  
Db 466 FVIANADNAGQYLARKADKVKSOEKKOLVVTTKDALDRAVAAYNALTAQOQTQOEKEKVDK 525  
QY 383 ---NFN-----DTNNVEMWETE---ANATEYTTGADGIIITGLKEGTYVLVEKKAPLGVN 431  
Db 526 AQAAYNAIAANAFEWADKNENNVKLVSDAQGRFEITGLLAGTYLLETKOPAGYA 585  
QY 432 LLDNSQKV-----ILGDGATDTNNDLNVPTVNNKGTETLPSTGGIGTTIFYIIG 483  
Db 586 LLTSRQKPEVATSYTSATQGIETAGSGK--DDATKVNKKITIPQITGGIITIFAVAG 643  
QY 484 ALLVIGAGIVLA 496  
Db 644 AAIM---GIAVYA 653

## RESULT 7

US-10-282-122A-60689  
; Sequence 60689, Application US/10282122A  
; Publication No. US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlssen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 60689

; LENGTH: 793

; TYPE: PRT

; ORGANISM: Listeria monocytogenes

US-10-282-122A-60689

Query Match 9.0%; Score 232; DB 12; Length 793;

Best Local Similarity 25.1%; Pred. No. 1.1e-09;

Matches 125; Conservative 66; Mismatches 189; Indels 118; Gaps 23;

QY 45 TYKAYKVF--DAEIDNANVSDNKGASLYLPQ-----GKEAEYK---- 82

Db 239 TYSADRVFTAGAPISSNFNS--ATSDGFSVALGNLTDLSVQISYITTTTDDGKSTQVDNTAK 297

QY 83 -ASTDP-----NSLFTTTTNG-----GRYVYTKD--TASANEIATW-----A 117

Db 298 LAGTDFVTTKQTSTWTPASGGGEGGGTTGTVLTKEADAKTKATLEGAEFKLVDSKGTVLQ 357

QY 118 KSISANTPTVSTVTESNNDGTEVINVSQYGYYSST-----VNNCAVINWTSVTENA 170

Db 358 ENITNAGSQLSIADLKPDYQLIETKAPTGYKLDTPVEFTIGENNOAI-----TVTKEN 413

QY 171 TIHEKNTDATWGDGKGK--TVDQKTVSVGDTVKYTYITYKNAVNVHYTEKV-----YQYV 222

Db 414 TLNTGSVELTKLDAATKATLAGATFELQDKENLTQDLKTDENGVLKVTDLVPGSYQFV 473

QY 223 IKDTMPSASVVDLNEGSEYVTTDGSNITTLTQSEKATGKYNLLENNNFTITIPWAA 282

Db 474 -ETSAPTGYKLDNSPVSEV--IAGETDQVVKVT-----KENTLEV-----SVELTK 518

QY 283 TNPPTGNTQNCANDDFYKGINITVTV-----TGVLK--SGAKPGSADLPEN---TNIATI 334

Db 519 LDSATKATLAGATFELQDKENLTQDLKTDENGVLKVTDLVPGSYQFVSAPTGYKLD 578

QY 335 NPNTS-----NDPQKQVTVRD---GQITIKKIDGSKASLOGAIFVLKNATQOFLNF 384

Db 579 NSPVSEFVAGETDQVVKVTENTLEVSGVSLTKDSATKATLAGATFELQDKENLTQ- 637

QY 385 NDTNNVWEGTEANATEYTTGADGIITITGLKEGTYVLVEKKAPLGNLDDNSQKVLGDG 444

Db 638 -----TGLTTDENGVLKVTDLVPGTYQFVETKAPIGYELDTTPVSEIIVAG 683

QY 445 ATD---TTNSDNLVNPT 459

Db 684 ETDPVVKVTENTLVPPT 701

## RESULT 8

US-09-864-761-36047

; Sequence 36047, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aecmica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30



```
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-35

Query Match      8.5%; Score 219; DB 14; Length 915;
Best Local Similarity 22.5%; Pred. No. 1.5e-08;
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;

QY 29 AAETGTTTVDQTKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQCKEAEYKASTD-- 86
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
309 AKDNSNLIGDSDAGNTDAKKV-----TFSNVSKDSKISASDHNVTLNSKVETSGDSTD 363
QY 87 -----FNSLFTTTTNGRTRYVTKDTASAN-----EIAATWAKSISANTTP 126
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
364 EDGNNNTGLTITAKNVTVNNNITSHKTVNITASENVTTKAGTTINATGTSVEVTAKTGD 423
QY 127 VSTVTESSNDGTEV-----INVSQYGYVYSSVTVNNGAVIMVSTPNATHEKNTDAT 180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
424 IKGGIESGNNVNITASGDTLNVSNITQGNVTVAASGAVTTTKGSTINATTGNANITTK 483
QY 181 WDGGGKTVDOKTYSGVGTVK-YTITYKNAVNYHGTEKYQYVVKDTPSASVVDLNEGS 239
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
484 TGEINGE-VKSAGNNVITASGNTLVNSNITQGNVTVTANSAL--TTTEGSTINATTGD 540
QY 240 YEVTTIDG-----SGNITTLTQSEKATGKYNLENNNFTITPWAATNPTGNT 290
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
541 ANITQTGNGKVESSGSVTLIATGOTLAVGNIS----GDTVTITADKGLTTQTSK 596
QY 291 ONGANDFFYKGI-----NTITVTYTVGL-----KSGAKPGSADLPENT 329
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 INGT-----KSVTSSQSGDISGTSVTSATSLTTQAGSKIEAKTGEANVTSAT 650
QY 330 NI--ATINPNTSNDPQKVTVRDQITIKKIDGSTKASLOGAIFVLKNATGQFLNFNDT 387
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
651 GTIGTISGNTVN-----VTANTDNLTIK--DGARIKATGGAVTL--TATGGTLTTETS 700
QY 388 NNVEWG-----TEANATEYTTGADGIITITGLK-----EGTYVLVEKKA 426
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
701 SDITSSNGQTLITAKDSSIAGSINAANVTLTNTGT--LTTVAGSKIEAASGTLVINAKDA 758
QY 427 PLGYNLLDNSQKVLGDGATDTTNSDNLNVNPTVENNKGTELPSTGGIGTTIFIIGAIL 486
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
759 QL-----DGA---ASGDHTVVNATNANGSGSVIATT-----SSRVNITGDLI 797
QY 487 VIGAGIVLVAR 497
Db : : : : :
798 TIN-GLNIISK 807

RESULT 11
US-10-193-764-37
; Sequence 37, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
```

```
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-37

Query Match      8.5%; Score 219; DB 14; Length 1222;
Best Local Similarity 22.5%; Pred. No. 2.3e-08;
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;

QY 29 AAETGTTTVDQTKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQCKEAEYKASTD-- 86
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
616 AKDNSNLIGDSDAGNTDAKKV-----TFSNVSKDSKISASDHNVTLNSKVETSGDSTD 670
QY 87 -----FNSLFTTTTNGRTRYVTKDTASAN-----EIAATWAKSISANTTP 126
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
671 EDGNNNTGLTITAKNVTVNNNITSHKTVNITASENVTTKAGTTINATGTSVEVTAKTGD 730
QY 127 VSTVTESSNDGTEV-----INVSQYGYVYSSVTVNNGAVIMVSTPNATHEKNTDAT 180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
731 IKGGIESGNNVNITASGDTLNVSNITQGNVTVAASGAVTTTKGSTINATTGNANITTK 790
QY 181 WDGGGKTVDOKTYSGVGTVK-YTITYKNAVNYHGTEKYQYVVKDTPSASVVDLNEGS 239
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
791 TGEINGE-VKSAGNNVITASGNTLVNSNITQGNVTVTANSAL--TTTEGSTINATTGD 847
QY 240 YEVTTIDG-----SGNITTLTQSEKATGKYNLENNNFTITPWAATNPTGNT 290
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
848 ANITQTGNGKVESSGSVTLIATGOTLAVGNIS----GDTVTITADKGLTTQTSK 903
QY 291 ONGANDFFYKGI-----NTITVTYTVGL-----KSGAKPGSADLPENT 329
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 INGT-----KSVTSSQSGDISGTSVTSATSLTTQAGSKIEAKTGEANVTSAT 957
QY 330 NI--ATINPNTSNDPQKVTVRDQITIKKIDGSTKASLOGAIFVLKNATGQFLNFNDT 387
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
958 GTIGTISGNTVN-----VTANTDNLTIK--DGARIKATGGAVTL--TATGGTLTTETS 1007
QY 388 NNVEWG-----TEANATEYTTGADGIITITGLK-----EGTYVLVEKKA 426
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1008 SDITSSNGQTLITAKDSSIAGSINAANVTLTNTGT--LTTVAGSKIEAASGTLVINAKDA 1065
QY 427 PLGYNLLDNSQKVLGDGATDTTNSDNLNVNPTVENNKGTELPSTGGIGTTIFIIGAIL 486
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1066 QL-----DGA---ASGDHTVVNATNANGSGSVIATT-----SSRVNITGDLI 1104
QY 487 VIGAGIVLVAR 497
Db : : : : :
1105 TIN-GLNIISK 1114

RESULT 12
US-10-193-764-34
; Sequence 34, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-34

Query Match      8.5%; Score 219; DB 14; Length 1228;
Best Local Similarity 22.5%; Pred. No. 2.4e-08;
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;

QY 29 AAGTGTITVQDTQKGYKAYKVFDAEIDNANVDSNKGASYLIPQGEAEYKASTD-- 86
Db 622 AKDNSLTIGDSDAGNTDAKV-----TFSNVKDSKISASDHNVTLSKSVTSGDTST 676

QY 87 -----FNSLFTTNGRVTYKDDTASAN-----BIATWAKSISANTTP 126
Db 677 EDGNNNTGLTITAKNVTVNNNITSHKTNITASENVTTKAGTTINATGSEVETAKTGD 736

QY 127 VSTVTSNNDGTVE-----INVSQYGYVSVSTVANGAVIMVTSVTPNATHEKNTDAT 180
Db 737 IKGIESNSGNVNITASGDTLVNSNITGQNVTVAAASGAVTTKSGSTINATGNANITTK 796

QY 181 WDGGGKTVQDKTYSYVGDIVK-YTITYKNAVNYHGTKEYVYQVYKIDMPSASVVDLNEGS 239
Db 797 TGEINGE-VKSASGNVNITASGNTLVNSNITGQNVTVTANSgai--TTTEGSTINATGSD 853

QY 240 YEVITIDG-----SGNITTLTQSEKATCKYNLLENNNFTITIPWAATNPTGNT 290
Db 854 ANITTQTNINGKVESSGSVTLIATGQTLAVGNIS-----GDTVTITADKGLTTQTSSK 909

QY 291 QNGANDDFYKGI-----NTITVYTGVL-----KSGAKPGSADLPENT 329
Db 910 INGT-----KSVTSSQSGDISGTTISGNTVSVATGSLTTQAGSKIEAKTGEANVTSAT 963

QY 330 NI--ATINPNTSNDPQGVKTVRDGQITIKIDGSTKASLQGAIFVLKNATGQFLNPDNT 387
Db 964 GTTGGTISGNTVN-----VTANTDNLTIK--DGAIRIKATGGAVTL--TATGGTLTTETS 1013

QY 388 NNVEWG-----TEANAEYTTGADGIITIGLK-----EGTVYVEKKA 426
Db 1014 SDITSSNGQTLTAKDSSAGSINAANVTNLTGT--LTTVAGSKIEAASGLVLVNAKOA 1071

QY 427 PLGYNLLDNSQKVLGDGADTTNSDNLVNPVTENNKGTELPSTGGIGTTIPYIIGAIL 486
Db 1072 QL-----DGA-----ASGDHTVNVNATNANGSGSVIATT-----SSRWNIITGDLI 1110

QY 487 VICAGIVLVAR 497
Db 1111 TIN-GLNIISK 1120

RESULT 13
US-10-369-493-13955
; Sequence 13955, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13955
; LENGTH: 1465
; TYPE: PRT
```

```
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13955

Query Match      8.5%; Score 218; DB 15; Length 1465;
Best Local Similarity 21.7%; Pred. No. 3.7e-08;
Matches 146; Conservative 61; Mismatches 207; Indels 260; Gaps 30;

QY 11 VASLAFGMAVSPVPIAFAAETGTTT-VQDTQKGATYKAYKVFDAEI-----DNAN 60
Db 146 LVTFTFSEAVTGFTNADLTIANGTILSAVSSDGGITGTGFTPTASITDTTNLISLDNTG 205

QY 61 VSD-----SNKDGASYLIPQGEAEYKASTDFN-----SLFTTTNGGRYTYTKKDT 107
Db 206 IADLAGNAGSGTTSANVVIDTVRPTATIVADSNLAAGETSLVITTFSEAVSGFTADL 265

QY 108 ASANEI-----ATWAKSISANTTPVSTVTSNN-----DGTEVI-----N 142
Db 266 TVANGTLSALSSSDGGITWTVTL-----TPTSNTITDNLITLDNTGVVDLAGNAGSGTTN 321

QY 143 VSOY-----GYVVSSTVANGAVIMVTS-- 165
Db 322 SNNYAIDTARPTATIVADNSLRIGETSLVITTFSEAITGLTNADLTIANGLTAVSSD 381

QY 166 -----VTPNATHEKNTDAT-----WGGGGKTVQDKTYS----- 195
Db 382 GGITWTATFTPTASI-----TDATNLITLNTGTIADLNGNAGSGTTDSNNYAIDTVRPTAT 437

QY 196 -----VGDVVKYTTYKNAVNYHGTKEYVYQVYKIDTM-----PS 229
Db 438 IVVANNURIGETSLVITTFSEAVS--GFTNADLTVANGTLSALSSDGGITWTATFTPS 495

QY 230 ASVVDL-----NEGSYEVITIDGSG-----NIT 252
Db 496 ASVDTTNLITLDNTGIADLAGNAGSGTTDSNNYAIDTQRPATIVADSNLSAGETSIV 555

QY 253 TLTQSEKATCKYNLLENNNFTIT-----IPWAATNPTGNTQNGANDDFYKGIN 304
Db 556 TIT-FSEAVTGFTNADLTVANGTILSAVSSDGGITWTATFTP-----NVGVD-----ATN 605

QY 305 TITVTYTVGLKSGAPGSADLPENTNIATINPNTSND-----PGQKTVVRDGOITI----- 356
Db 606 LITLANTGI-----ADLSGNTSGTTSNNYSIDTVLTATIVADNALKIGETSL 656

QY 357 -----KKIDGSTKASLQGAIFVLKNATGQFLNFDNTNNVWGTEBANAFTYTTGADGIIT 410
Db 657 VTITFSEAVTGFTNADL-----TIANGTILSAVSSD--GGITWTATFTPTTSITDATNLIT 710

QY 411 ITGLKEGYLYVEKAPLGNLLDNS-QKVILGDGATDTTNSDNLVNPVTENNKGTELP 469
Db 711 -----LDNSGVQNLSGNAGSGTTDSNNYAID-TVRPATMLVA 747

QY 470 STG-GIGTTIFVII 482
Db 748 DTALGIGQTLVTI 761

RESULT 14
US-10-282-122A-73634
; Sequence 73634, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
;; FILE REFERENCE: ELITRA.034A  
;; CURRENT APPLICATION NUMBER: US/10/282,122A  
;; CURRENT FILING DATE: 2003-02-20  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: Patent in version 3.1  
;; SEQ ID NO 73634  
;; LENGTH: 645  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
;; FEATURE:  
;; NAME/KEY: MISC FEATURE  
;; LOCATION: (569)..(569)  
;; OTHER INFORMATION: X=any amino acid  
;; FEATURE:  
;; NAME/KEY: MISC FEATURE  
;; LOCATION: (586)..(586)  
;; OTHER INFORMATION: X=any amino acid  
;; FEATURE:  
;; NAME/KEY: MISC FEATURE  
;; LOCATION: (617)..(617)  
;; OTHER INFORMATION: X=any amino acid  
;; FEATURE:  
;; NAME/KEY: MISC FEATURE  
;; LOCATION: (641)..(641)  
;; OTHER INFORMATION: X=any amino acid  
US-10-282-122A-73634

Query Match 8.2%; Score 212; DB 12; Length 645;  
Best Local Similarity 23.3%; Pred. No. 3.4e-08;  
Matches 121; Conservative 57; Mismatches 194; Indels 148; Gaps 24;  
QY 34 TITVQDTQKATYKAYKVFDAIDN-----ANVSDSKDGASYLI 73  
DB 26 TTTVDGKNGKGRAPKIKVEDITSPSIRRDAAATPRNGIRVVYDVNDNGV-- 83  
QY 74 PQCKEAYKASTDFNSL-----FTTTNGRTYVTKDTSANETATWAKSISANT 124  
DB 84 DGVGVKLSKIYNGIDRGDSAPTITTKDNGDGHHTTVQNPDSSESTTVKDGKOGK 143  
QY 125 TPVSTVTEENNNGTEVINVSQYGYVSVTVANGAVIMVTSVTPNATIEKN-----TDA 179  
DB 144 TANITITTE-NPDGSHITVTPNPDGSPKETVKNKGK----DGKTPKVEVTDNNDGHTVTKV 198  
QY 180 TWGDGG-----GKTVQKIVSGDVTVKYIITYQAVNYHGTEKYQYVVKD-- 225  
DB 199 TDGNGVNTNAILKDGKDGKAATATTENPDG-SHTVITNP---DGTKN--EFVVKNGRD 252  
QY 226 ----TPMSASVDNLNGSEVETITDGSNITTLTQGSERATGKYNLLENNNNFTTIPWA 281  
DB 253 GVDGRTPTASVRDNGDGHSHIIVITNPEG-VTTETTVROGSKSPKVTITDE----- 300

QY 282 ATNTPTGNTQNGANDDPFYKGINTITVTYTGVLKSGAKPGSADLPENTNIATINPNTSND 341  
DB 301 -----QNGTHKISVLNGDGTITET---IIKDGKSPVATVRDNDQDGTYYIRVENGNG 348  
QY 342 DPGQKVTVRDGO-----ITIKIDGSTRASLQGAIFVLKNATGQFLNFND 386  
DB 349 TVSE-TTVRDGKSPAKVVDNGDGHHTITVNSDGIITTT-----TVRDGREGKLEVID 401  
QY 387 TNNVEWGTENATEYTTGADGIITITGLKEGTYIYVLEKAPLGNLLDNSQ----- 437  
DB 402 NN-----DGSHTIKVTGADGKGTITTFD-----KSPKA-NIVDNGDGHHTLTIVD 447  
QY 438 -----KVILGDGATDTTNSDNLVNPV---ENKKGTEL 468  
DB 448 SDGREYKSIKDG-KDGKDS-----VSPTVTVKNNDGTHV 482  
RESULT 15  
US-10-282-122A-74083  
;; Sequence 74083, Application US/10282122A  
;; Publication No. US20040029129A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, Liangsu  
;; APPLICANT: Zamudio, Carlos  
;; APPLICANT: Malone, Cheryl  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Kari  
;; APPLICANT: Zyskind, Judith  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John  
;; APPLICANT: Carr, Grant  
;; APPLICANT: Yamamoto, Robert  
;; APPLICANT: Forsyth, R.  
;; APPLICANT: Xu, H.  
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
;; FILE REFERENCE: ELITRA.034A  
;; CURRENT APPLICATION NUMBER: US/10/282,122A  
;; CURRENT FILING DATE: 2003-02-20  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: Patent in version 3.1  
;; SEQ ID NO 74083  
;; LENGTH: 2551  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
US-10-282-122A-74083

Query Match 8.2%; Score 212; DB 12; Length 2551;  
Best Local Similarity 23.3%; Pred. No. 2.5e-07;  
Matches 121; Conservative 57; Mismatches 194; Indels 148; Gaps 24;  
QY 34 TITVQDTQKATYKAYKVFDAIDN-----ANVSDSKDGASYLI 73

Db 1891 TFTVKGKNGKGRAPKIKVEDITSPSIRRRDAAAATPRNGIRVTVVDDVNDNGVY-- 1948  
QY 74 PQKEAEYKASTDFNSL-----FTTTNGGRTVVTVKOTASANEIATWAKSIANT 124  
Db 1949 DEGVKVLNSKDIYNGIDGRDGSAPITTKDNGDGHHTITVQNPQDSESTTVVKDGKDGK 2008  
QY 125 TPVSTVTESSNDGTEVINVSQYGYVVSFVNNGAVIMVTSVTPNATIHKN-----TDA 179  
Db 2009 TANITITTE-NPDGSHITITVNPDSGTKETVVKNGK-----DGKTPKVEVTDNNDGTHVKV 2063  
QY 180 TWGDGG-----GKTVDKQTVSGVTVKYITTYKNAVNYHGTETKVYQVVKD--- 225  
Db 2064 TDGSGNVTNIIKDGKDGKAATATTENPDG-SHTVTITNP-----DGTKN--EFVWKNGRD 2117  
QY 226 ---TWPSASVLDNLSGSEVITITDGSNITTTTQSGEKATGKYNLLENNNFTITIPWA 281  
Db 2118 GVDGRPTASVRNGDGSHTIVITNPEG-VTETTVRDGKSPKVTITDE----- 2165  
QY 282 ATNPTGTONTGANDFFYKGINITITVYGVLSKAGKPSADLPENTNIATINPNTSND 341  
Db 2166 -----QNGTHKISVLNGDGTETET---IIKDGKSPVATVRDNDGTYTIRVENGG 2213  
QY 342 DPQOKVTVRDQ-----IITIKIDGSTKASLOGAIFVLKNATQOFLNFD 386  
Db 2214 TVSE-TTVRDGKSPTAKVDNDGSHHTITVNSDGIITTT-----TVRDGREPKLEVID 2266  
QY 387 TNNVWGTENATEYTTGADGIITITGLKEGTYLVKKGAPLGYNLLDNSQ----- 437  
Db 2267 NN-----DGSHTIKVTGADGKGTITITPDG-----KSPKA-NIVDNGDGHHTITIVD 2312  
QY 438 -----KVILGDGATITTSNDLLNPTV---ENNKGTGL 468  
Db 2313 SDGREYKSIKDG-KDGKDS---VSPTVTVKNNNDGTHV 2347  
  
RESULT 16  
US-10-661-809-11  
; Sequence 11, Application US/10661809  
; Publication No. US20040101919A1  
; GENERAL INFORMATION:  
; APPLICANT: HOOK, Magnus  
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR  
; FILE OF INVENTION: POSITIVE BACTERIA  
; FILE REFERENCE: P07741US01/BAS  
; CURRENT APPLICATION NUMBER: US/10/661,809  
; PRIOR FILING DATE: 2003-09-15  
; PRIOR APPLICATION NUMBER: 60/410303  
; PRIOR FILING DATE: 2002-09-13  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 11  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-661-809-11  
  
Query Match 8.1%; Score 208.5; DB 16; Length 476;  
Best Local Similarity 26.5%; Pred. No. 4.2e-08;  
Matches 85; Conservative 49; Mismatches 124; Indels 63; Gaps 14;  
  
QY 187 KTV--DQKTVSGDVTKYITTYKNAVNYHGTETKVYQVVKDTPSASVLDNLSGSEVIT 244  
Db 192 KTVLDKQGFNGOGEHINYQLTTQIPANILG-----YQ-----EFLR 227  
QY 245 TDGSGNITTLTQGS--EKATGK-----YNLLENNNFTITIPWAATNPTGTONTGANDD 297  
Db 228 SDKADITLTLPLSIEVKVAGKVTTCYITLTQKHGFTL-----DPSIKDLQNFAN-- 278  
QY 298 FFYKGINITITVYGVLSKAGKPSADLPENTNIATINPNTSNDPQKQVTVRQDGIITK 357  
Db 279 -----QTMVTSVQMRLEKTAEPDTAINNEGQLV-----TDKHTLTKRATVRTGKGSFV 326  
QY 358 KIDG-STKASLOGAIFVLKNATQOFLNFDNNTNVEWGTE-ANATEYTTGADGIITITGLK 415

Db 327 KVDSENAKITLPEAVFIVKNOAGEYLN--ETANGYRWOKEKALAKKFTSNQAGEFSVKGLK 385  
QY 416 EGYTVLVKPKAPLGYNLLDNSQKVLGDGATDTTNSDNLNVNPTVENNKGTG--LPSTG 472  
Db 386 DGQYFLEESAPKGYLLNQTEIPFTVGNKNSYATNGQRTAPLH--VINKVKESGFLPKTN 443  
QY 473 GIGTTIFVFIIGAILVIGAGIV 493  
Db 444 E-ERSIWLTIAGLLIIGMWVI 463  
  
RESULT 17  
US-09-769-787-126  
; Sequence 126, Application US/09769787  
; Publication No. US20030091577A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe PG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129WO  
; CURRENT APPLICATION NUMBER: US/09/769,787  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 126  
; LENGTH: 886  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-769-787-126  
  
Query Match 7.8%; Score 200.5; DB 10; Length 886;  
Best Local Similarity 20.6%; Pred. No. 4.5e-07;  
Matches 107; Conservative 69; Mismatches 164; Indels 179; Gaps 24;  
  
QY 69 ASYLIPQKGAEYKASTDFNSLFTTTNGGRTVVTVKOTASANEIATWAKSI-----SAN 123  
Db 456 AAFPVKPEKYSKAG--YAVIGDPINGGYIWL-----NWRESILAYPFNSN 501  
QY 124 TTPVSTVTESSNDGTEVINVSQYGYVVSFVNNGAVIMVTSVTPNATIHKNATDWDG 183  
Db 502 TAKIT-----NHGDPTR-----WYNGNIAPDGYDVFTVGIGING---DPGTDEATAT 546  
QY 184 GGGKTVDOQT---YSVGDIVK-----YITTYKNAVNYHGTETKVYQVVKDTPSAS 231  
Db 547 SFMQSISKSPENYTNVDTTKILEQLNRYFHTIVTEKKSIENT-----ITDPMGE 597  
QY 232 VVDLNEGS-----YEVTITDGS-----GNITTLTQSGE---AT 262  
Db 598 LIDLQGTGDRFPDADYTLTANDGSELENGQAVGQPDGGLLKAKVLYDTEKRIVT 657  
QY 263 GKNLLENNNFTITIPWAATNPTGTONTGANDDP-----FYKGINITITVYGVLSKGA 318  
Db 658 GLYLGTDKVTILTYN-----RLNDEFVSNKFY----- 685  
QY 319 KPGSADLPENTNIATNP-----NTSNDPQKQVTVRD-----GQITIKK 358  
Db 686 -----DTNGRTLLHPKEVBQNTVRDPPIPK--IRDVRKYPEITISKEKLGDIETIK 735  
QY 359 IDGSTKASLOGAIFVLKNATQOFLNFDNNTNVEWGTEANAT--EYTTGADGIITITGLKE 416  
Db 736 VNKNDKKPLRGAVFSLOK-----QHPDYPDIYGAIDQNGTYQNVRTGDKGLTFKNLSD 789  
QY 417 GTYLVKPKAPLGYNLLDNSQ-----KVILGD-----GATDTTNSDNLNPT 459  
Db 790 GKRYLPENSEPAGYKPVQNKPIVAFQIVNGEVRDVTISVPODIPAGYEFNDKHYITNBP 849  
QY 460 VENNKGTELPSTGGIGTTIFVFIIGAILVIGAGIVLVARR 498

Db 850 IPPKR--EYPRTGGIGMLPFYLGCMMM--GGVLLYTRK 884

**RESULT 18**

```

US-10-661-809-17
; Sequence 17, Application US/10661809
; Publication No. US2004010191A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD
; TITLE OF INVENTION: POSITIVE BACTERIA
; FILE REFERENCE: P07741US01/BAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 473
; TYPE: PRT
; ORGANISM: staphylococcus epidermidis
US-10-661-809-17

```

Query Match	7.7%	Score 199;	DB 16;	Length 473;
Best Local Similarity	21.5%;	Pred. No. 2.4e-07;		
Matches 118;	Conservative	84;	Mismatches 216;	
			Indels 130;	Gaps 23;

QY 1 MKKKMIQSLLVASLAFGMAVSPVTPIAFAAETGTTVQ-----DTQ 41

D6 MKKI GWT.SMCI.FI.I.I.EKPA TOV-----A TETEMVOIT.T.HKI.I.FPNGOI.PKNHPNDGO 55

QY 42 KGATYKAYKVFD AEIDNANVSDSNKDGASYLIPOGKEAEYKASTDENSLFTTTTNGGRTY 101

nh  
56 EKALITVB-----CI NCIMTEVI-----YDMMNGEVVI---PBYCVTVI 80

OV 102 VTKKDTASANEIATWAKSISANTTPVSTVTESNNDGTEVINVS-----QYGYVYVSST 154

[illegible]

155 V-----NNGAVIMVTSVTPN---ATTH--EKNTDATWGNGGGKTV---DOKTVSYGD 198

[illegible]

199 TVKVITTYKNAVYHCTEKKWQVVKDTPPSASWVDINEGSVEVTTITDCSGNITTTITGCS 258

[illegible]

250 BYATCKVNI I PENANISIT - TI DIA LATATDTANTONCANDREBYUCNTNITITITUTUTUHV 216

[illegible]

30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530  
 531  
 532  
 533  
 534  
 535  
 536  
 537  
 538  
 539  
 540  
 541  
 542  
 543  
 544  
 545  
 546  
 547  
 548  
 549  
 5

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]

Variable	Mean	SD	Min	Max
Age	35.2	12.5	18	65
Gender	0.45	0.50	0	1
Marital Status	0.60	0.49	0	1
Education	12.8	2.1	9	16
Income	45000	15000	20000	80000
Health Status	0.75	0.43	0	1
Employment Status	0.80	0.41	0	1
Life Satisfaction	4.2	1.5	1	7
Stress Level	3.8	1.2	1	6
Resilience Score	5.5	1.8	2	8
Optimism Index	6.0	1.0	3	8
Gratitude Level	5.8	1.2	3	8
Forgiveness Score	6.2	1.5	3	8
Empathy Index	6.5	1.0	3	8
Emotional Stability	6.8	1.2	3	8
Life Purpose	7.0	1.5	3	8
Meaning in Life	7.2	1.0	3	8
Existential Well-being	7.5	1.2	3	8
Transcendental Experience	7.8	1.5	3	8
Spiritual Growth	8.0	1.0	3	8
Inner Peace	8.2	1.2	3	8
Self-Actualization	8.5	1.5	3	8
Peak Experience	8.8	1.0	3	8
Flow State Frequency	9.0	1.2	3	8
Authenticity Score	9.2	1.5	3	8
Personal Growth	9.5	1.0	3	8
Life Satisfaction (Revised)	9.8	1.2	3	8
Overall Well-being	10.0	1.5	3	8

Q	A
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

## RESULT 19

US-10-282-122A-45795  
; Sequence 45795, Application US/10282122A



```

Db      362  ----SLTKGAVETKVVDDICATLEGAVFKIVD-----WNGNDVR-----TDLT 402
Qy      404  GAGGIITITGLKEGYTLVEKKAPLGVNLIDN--SOKVILGDGATDTTNSDMLLVNPTVE 461
         ||:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      403  DKDKGISVSDLRPGDYQFIETKAPHYDLNQNPINFTEVKESQTATASVTATNSLTKGAVE 462
Qy      462  NNNKETPLSPGGIGTIIFYIIGAILVI 488
Db      463  LMKVDDID----GTT---LEGAIFKI 481

RESULT 20
US-09-946-374-243
; Sequence 243, Application US/09946374
; Publication No. US20030073129A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PICI
; CURRENT APPLICATION NUMBER: US/09/946,374
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792

```

PRIOR APPLICATION NUMBER: 60/102207  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102240  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102307  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102330  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102331  
PRIOR FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: 60/102484  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 60/102487  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 60/102570  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 60/102571  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 60/102684  
PRIOR FILING DATE: 1998-10-01  
PRIOR APPLICATION NUMBER: 60/102687  
PRIOR FILING DATE: 1998-10-01  
PRIOR APPLICATION NUMBER: 60/102965  
PRIOR FILING DATE: 1998-10-02  
PRIOR APPLICATION NUMBER: 60/103258  
PRIOR FILING DATE: 1998-10-06  
PRIOR APPLICATION NUMBER: 60/103314  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103315  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103328  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103395  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103396  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103401  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: 60/103449  
PRIOR FILING DATE: 1998-10-06  
PRIOR APPLICATION NUMBER: 60/103633  
PRIOR FILING DATE: 1998-10-08  
PRIOR APPLICATION NUMBER: 60/103678  
PRIOR FILING DATE: 1998-10-08  
PRIOR APPLICATION NUMBER: 60/103679  
PRIOR FILING DATE: 1998-10-08  
PRIOR APPLICATION NUMBER: 60/103711  
PRIOR FILING DATE: 1998-10-08  
PRIOR APPLICATION NUMBER: 60/104257  
PRIOR FILING DATE: 1998-10-14  
PRIOR APPLICATION NUMBER: 60/104987  
PRIOR FILING DATE: 1998-10-20  
PRIOR APPLICATION NUMBER: 60/105000  
PRIOR FILING DATE: 1998-10-20  
PRIOR APPLICATION NUMBER: 60/105002  
PRIOR FILING DATE: 1998-10-20  
PRIOR APPLICATION NUMBER: 60/105104  
PRIOR FILING DATE: 1998-10-21  
PRIOR APPLICATION NUMBER: 60/105169  
PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: 60/105266  
PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: 60/105693  
PRIOR FILING DATE: 1998-10-26  
PRIOR APPLICATION NUMBER: 60/105694  
PRIOR FILING DATE: 1998-10-26  
PRIOR APPLICATION NUMBER: 60/105807

Query Match 7.7%; Score 198; DB 10; Length 596;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

14 LAFGM-----AVSPVTPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNANVSDSNKDG 68

QY

DB 10 LMFGLLLHLEAATNSNETSTANTGSSVSSGASTATNSGSSVTSGSVSTATISGSSVTS 69  
QY 69 ASYLIPQCKEAEYKA---STDFNSLFTTTTNGRTVYVKDTASANEIATMAKSIANTT 125  
DB 70 NGVSVITNSEPHSTSSGISTATNSSEFSTASS-GSIATNSSESTSSGASTATN-SEST 127  
QY 126 PVSTVTESNNDGTEVINVSQYGYVSVSVNANGAVIMVTSVTPNATIIHEKNTDATWGDGG 185  
DB 128 PSSGASTVTSNGSSVTSSG-----ASTATNSSESTVSSRASTATNSSESTLSS-GAST 179  
QY 186 GKTVDQKTSYSGDVTKYITTYKNAVNYHGTEKYVYVVKDTPMSASVVDLNEGSYEVTIT 245  
DB 180 ATNSDSSTSSGAS-----TATNSSESS-----TTSSGASTATNSSESS---TVS 219  
QY 246 DSGNIT-----TLTQSEKATGKYNLLENNFTITPWAATNPTGTQNGANDDPFY 300  
DB 220 SRASATNSSESTSSGASTATNSERSTTSNAGT-----ATNSSESTSSGA----- 267  
QY 301 KGINTIVTYTGVLSKAGKPGSADLPENTNIATINPNTSNDPPGQKVTVRDQGITIKKID 360  
DB 268 ---STATNSDSSTVSSGASTAT-----NSESSTSSGASTATNSSESTSSGASTATNSD 319  
QY 361 GSTKASLOGAIFVLKNATGQFLNFNDTNVWEGTEANATEYTTGADGIITITGLKEGIYY 420  
DB 320 SSTTSS--GA-----GTATN-SESSTVSSGISTVTNSSEST-- 352  
QY 421 LVEKKAPLGYNLLDNSQKVLGDGATDTNTSNDLLNVPTVENNKGTLPST-GGIGT 476  
DB 353 -----PSSGANTATNSSESTSSGANTATNSSESTVSSGASTATNSSESTSSGVST 404

## RESULT 21

US-10-206-915-310

; Sequence 310, Application US/10206915

; Publication No. US20040029221A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P34301C513

; CURRENT APPLICATION NUMBER: US/10/206,915

; PRIOR FILING DATE: 2002-07-26

; PRIOR APPLICATION NUMBER: 10/052586

; PRIOR FILING DATE: 2002-01-15

; PRIOR APPLICATION NUMBER: 60/059263

; PRIOR FILING DATE: 1997-09-18

; PRIOR APPLICATION NUMBER: 60/059266

; PRIOR FILING DATE: 1997-09-18

; PRIOR APPLICATION NUMBER: 60/062250

; PRIOR FILING DATE: 1997-10-17

; PRIOR APPLICATION NUMBER: 60/063120

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: 60/063121

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: 60/063486

; PRIOR FILING DATE: 1997-10-21

; PRIOR APPLICATION NUMBER: 60/063540

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/063541

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/063544

; PRIOR FILING DATE: 1997-10-28





Db 268 ---STATNSDSSTVSGASTAT-----NSBSSTSSGASTATNSBSSTSSGASTATNSD 319  
QY 361 GSTKASIQAGI FVLKATGQFLNFNDTNVVEWTEANATYTTGADGIITITGLKEGTY 420  
Db 320 SSTSS--GA-----GTATN-SESSTVSGISTVTSNSESST-- 352  
QY 421 LVEKKAPLGYNLLDNSOKVILGDGATDTTNSDMLLVNPTVNNKGTLPST--GGIGT 476  
Db 353 -----PSSGANTATNSBSSTSSGANTATNSBSSTSSGASTATNSBSSTSSGAST 404

RESULT 25  
US-10-208-024-310  
; Sequence 310, Application US/10208024  
; Publication No. US20040048335A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C538  
; CURRENT APPLICATION NUMBER: US/10/208, 024  
; CURRENT FILING DATE: 2002-07-29  
; PRIOR APPLICATION NUMBER: 10/052586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; PRIOR FILING DATE: 1997-10-28  
; Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-208-024-310

Query Match 7.7%; Score 198; DB 12; Length 596;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

QY 14 LAFGM-----AVSPVTPIAFAETGTTVQDTQKGATYKAYKVFDAEIDNANVDSNKG 68  
Db 10 LMFGLLLHLEAAVNSNETSANTSGSVSSGASTATNSGSSVTSSTSSGASTATNSGSSVTS 69  
QY 69 ASYLI PQGEAEYKA---STFNSLETTTNGRTYVTKKDTASANEIATWAKSISANTT 125  
Db 70 NGVSIYVNSEFHTTSSGISATNSSESTASS-GISATNSSESSTSSGASTATN-SEST 127  
QY 126 PVSVTESNNDGTEVINVSQYGYVSVSVNNGAVIMVTSVTPNATIHKNTDATWGDGG 185

Db 128 PSSGASTVNSGSSVTSSG-----ASTATNSSESTVSSRSTATNSSESTLSS-GAST 179  
QY 186 GKTVDQKTVSGDVTVKYVITYKNANVHGTEKYQYVVKDTPMPSASVDLNEGSYEVTIT 245  
Db 180 ATNSDSSTSSGAS-----TATNSSES-----TTSSGASTATNSSES---TVS 219  
QY 246 DSGNIT-----TLTQSEKATGKYNLLSENNFTITIPWAATNTPTGNTQKANDDPFY 300  
Db 220 SRATATNSSESTSSGASTATNSBSRTTSGAGT-----ATNSSESTSSGA----- 267  
QY 301 KGINTITVTYTCVLKSGAKPGSADLPENTNIATINPNTSNDPQKVTVRDQGITIKKID 360  
Db 268 ---STATNSDSSTVSGASTAT-----NSBSSTSSGASTATNSBSSTSSGASTATNSD 319  
QY 361 GSTKASIQAGI FVLKATGQFLNFNDTNVVEWTEANATYTTGADGIITITGLKEGTY 420  
Db 320 SSTSS--GA-----GTATN-SESSTVSGISTVTSNSESST-- 352  
QY 421 LVEKKAPLGYNLLDNSOKVILGDGATDTTNSDMLLVNPTVNNKGTLPST--GGIGT 476  
Db 353 -----PSSGANTATNSBSSTSSGANTATNSBSSTSSGASTATNSBSSTSSGAST 404

RESULT 26  
US-10-201-853-310  
; Sequence 310, Application US/10201853  
; Publication No. US20040053358A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C465  
; CURRENT APPLICATION NUMBER: US/10/201, 853  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: 10/052586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; PRIOR FILING DATE: 1997-10-28  
; Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-201-853-310



Db 180 ATNSDSSTSSGAS-----TATNSSS-----TSSGASTATNSSS---TVS 219  
Qy 246 DSGNIT-----TLTGSEKATGYNLLLENNFTTIPWAAATNTPTGNTONGANDFFY 300  
Db 220 SRATATNSSSTSSGASTATNSSERTTSGAGT-----ATNSSSTSSGA----- 267  
Qy 301 KGINTITVTYGVLKSGAKPGSADLPENTNIATINPTNSNDPQGVTVRDGQITIKKID 360  
Db 268 ---STATNSDSSTVSSGASTAT-----NSESSTSSGASTATNSSSTSSGASTATNSD 319  
Qy 361 GSTKASLOGAIFVLKNATQFLNFNDTNVWGTGANATEYTTGADGIITITGLKEGTY 420  
Db 320 SSTSS--GA-----GTATN-SESSSTVSSGISTVTNSSST-- 352  
Qy 421 LVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLVNPVTENNKGTELPST--GGIGT 476  
Db 353 -----PSSGANTATNSSSTSSGANTATNSSSTVSSGASTATNSSSTSSGVST 404

RESULT 29  
US-10-063-513-100  
; Sequence 100, Application US/10063513  
; Publication No. US20030018172A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063, 513  
; CURRENT FILING DATE: 2002-05-01  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-513-100

Query Match 7.7%; Score 198; DB 12; Length 596;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

Qy 14 LAFGM-----AVSPVTPPIAFAAETGTTITVQDTQKATYKAYKVFDAEIDNANVSDNKG 68  
Db 10 LMFGLLLHLEATNSNETSANTGSSVSSGASTATNSSSTVSSGVSTATISGSVTS 69  
Qy 69 ASYLIPOGKEAEYKA---STDFNSLFTTTTNGRTYVTKDPTASANEIATWAKSISANTT 125  
Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS--GISIATNSSSTSSGASTATN-SEST 127  
Qy 126 PVSTVTESNNDGTEVINVSQYGYVVSSTVNGAVIMVTSVTPNATIHEKNTDATWGDGG 185  
Db 128 PSSGASTVTVNGSSVTSSG-----ASTATNSSSTVSSRASTATNSSSTLSS--GAST 179  
Qy 186 GKTVDOKTVSGDVTYKVIITYKNVNYHGTEKYQYVVIKDTMPSASVVDLNEGSYEVTIT 245  
Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS--GISIATNSSSTSSGASTATN-SEST 127  
Qy 126 PVSTVTESNNDGTEVINVSQYGYVVSSTVNGAVIMVTSVTPNATIHEKNTDATWGDGG 185  
Db 128 PSSGASTVTVNGSSVTSSG-----ASTATNSSSTVSSRASTATNSSSTLSS--GAST 179  
Qy 186 GKTVDOKTVSGDVTYKVIITYKNVNYHGTEKYQYVVIKDTMPSASVVDLNEGSYEVTIT 245  
Db 180 ATNSDSSTSSGAS-----TATNSSS-----TSSGASTATNSSS---TVS 219  
Qy 246 DSGNIT-----TLTGSEKATGYNLLLENNFTTIPWAAATNTPTGNTONGANDFFY 300  
Db 220 SRATATNSSSTSSGASTATNSSERTTSGAGT-----ATNSSSTSSGA----- 267  
Qy 301 KGINTITVTYGVLKSGAKPGSADLPENTNIATINPTNSNDPQGVTVRDGQITIKKID 360  
Db 268 ---STATNSDSSTVSSGASTAT-----NSESSTSSGASTATNSSSTSSGASTATNSD 319  
Qy 361 GSTKASLOGAIFVLKNATQFLNFNDTNVWGTGANATEYTTGADGIITITGLKEGTY 420  
Db 320 SSTSS--GA-----GTATN-SESSSTVSSGISTVTNSSST-- 352  
Qy 421 LVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLVNPVTENNKGTELPST--GGIGT 476  
Db 268 ---STATNSDSSTVSSGASTAT-----NSESSTSSGASTATNSSSTSSGASTATNSD 319

Qy 361 GSTKASLOGAIFVLKNATQFLNFNDTNVWGTGANATEYTTGADGIITITGLKEGTY 420  
Db 320 SSTSS--GA-----GTATN-SESSSTVSSGISTVTNSSST-- 352  
Qy 421 LVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLVNPVTENNKGTELPST--GGIGT 476  
Db 353 -----PSSGANTATNSSSTSSGANTATNSSSTVSSGASTATNSSSTSSGVST 404

RESULT 30  
US-10-063-515-100  
; Sequence 100, Application US/10063515  
; Publication No. US20030018173A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063, 515  
; CURRENT FILING DATE: 2002-05-01  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-515-100

Query Match 7.7%; Score 198; DB 12; Length 596;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

Qy 14 LAFGM-----AVSPVTPPIAFAAETGTTITVQDTQKATYKAYKVFDAEIDNANVSDNKG 68  
Db 10 LMFGLLLHLEATNSNETSANTGSSVSSGASTATNSSSTVSSGVSTATISGSVTS 69  
Qy 69 ASYLIPOGKEAEYKA---STDFNSLFTTTTNGRTYVTKDPTASANEIATWAKSISANTT 125  
Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS--GISIATNSSSTSSGASTATN-SEST 127  
Qy 126 PVSTVTESNNDGTEVINVSQYGYVVSSTVNGAVIMVTSVTPNATIHEKNTDATWGDGG 185  
Db 128 PSSGASTVTVNGSSVTSSG-----ASTATNSSSTVSSRASTATNSSSTLSS--GAST 179  
Qy 186 GKTVDOKTVSGDVTYKVIITYKNVNYHGTEKYQYVVIKDTMPSASVVDLNEGSYEVTIT 245  
Db 180 ATNSDSSTSSGAS-----TATNSSS-----TSSGASTATNSSS---TVS 219  
Qy 246 DSGNIT-----TLTGSEKATGYNLLLENNFTTIPWAAATNTPTGNTONGANDFFY 300  
Db 220 SRATATNSSSTSSGASTATNSSERTTSGAGT-----ATNSSSTSSGA----- 267  
Qy 301 KGINTITVTYGVLKSGAKPGSADLPENTNIATINPTNSNDPQGVTVRDGQITIKKID 360  
Db 268 ---STATNSDSSTVSSGASTAT-----NSESSTSSGASTATNSSSTSSGASTATNSD 319  
Qy 361 GSTKASLOGAIFVLKNATQFLNFNDTNVWGTGANATEYTTGADGIITITGLKEGTY 420  
Db 320 SSTSS--GA-----GTATN-SESSSTVSSGISTVTNSSST-- 352  
Qy 421 LVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLVNPVTENNKGTELPST--GGIGT 476  
Db 353 -----PSSGANTATNSSSTSSGANTATNSSSTVSSGASTATNSSSTSSGVST 404



Search completed: July 22, 2004, 10:34:40  
Job time : 49 secs

---

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 22:18:23 ; Search time 6024 Seconds  
(without alignments)  
10857.344 Million cell updates/sec

Title: US-10-009-254-1  
Perfect score: 1509  
Sequence: 1 atgaaaagaagaattcatca.....gtcgtcgtttacgtttcttaa 1509

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 90 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sv.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pi.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1509	100.0	1509	1	AF485279	AF485279 Streptococcus
2	1509	100.0	1509	6	AX476887	AX476887 Sequence
3	92.8	6.1	216959	10	AC116998	AC116998 Mus muscu
4	89.6	5.9	171104	2	BX649545	BX649545 Danio rer
5	87.2	5.8	170221	10	AC127421	AC127421 Mus muscu
6	86.4	5.7	213544	2	BX510939	BX510939 Danio rer
c 7	86	5.7	116807	2	BX890614	BX890614 Danio rer
c 8	85.2	5.6	146275	2	AL935272	AL935272 Danio rer
c 9	84	5.6	198743	2	BX530070	BX530070 Danio rer
c 10	83.4	5.5	155204	2	AC007926	AC007926 Trypanoso
c 11	82.4	5.5	94534	5	AL929250	AL929250 Zebrafish
c 12	80.6	5.3	110000	2	PFMAL13_07	Continuation (8 of
c 13	80.4	5.3	225581	2	BX537105	BX537105 Danio rer
c 14	80	5.3	10115	5	AF397467	AF397467 Ictalurus
c 15	79.8	5.3	308092	3	AY333070	AY333070 Drosophila
c 16	78.6	5.2	224391	2	AC137124	AC137124 Mus muscu
c 17	78	5.2	164399	3	PFMAL3P6	Z98551 Plasmodium
c 18	77.8	5.2	154160	2	BX571709	BX571709 Danio rer
c 19	77.8	5.2	177595	2	AL953867	AL953867 Danio rer
c 20	77.4	5.1	200634	5	BX004990	BX004990 Zebrafish
c 21	77.2	5.1	176010	5	AL935062	AL935062 Zebrafish
c 22	75.6	5.0	151341	5	AL929536	AL929536 Zebrafish
c 23	75.4	5.0	22398	5	FRU21723	AJ271723 Fugu rubr
c 24	75.2	5.0	245292	2	BX088589	BX088589 Danio rer
c 25	75	5.0	183357	2	BX569779	BX569779 Danio rer
c 26	75	5.0	183357	2	BX569779	BX569779 Danio rer
c 27	74	4.9	199698	2	BX004991	BX004991 Danio rer
c 28	73.8	4.9	250029	3	AF014820	AF014820 Plasmodiu
c 29	73	4.8	666	8	AF413050S2	AF413051 Zea mays
c 30	73	4.8	199347	2	BX321884	BX321884 Danio rer
c 31	72.8	4.8	146907	2	AC141031	AC141031 Rattus no
c 32	72.6	4.8	173134	5	AL935282	AL935282 Zebrafish
c 33	72.4	4.8	157033	2	BX323881	BX323881 Danio rer
c 34	72.4	4.8	179553	2	AC024253	AC024253 Homo sapi
c 35	72.2	4.8	153751	3	AC116551	AC116551 Dictyoste
c 36	72.2	4.8	180668	2	AC020857	AC020857 Mus muscu
c 37	70.6	4.7	146415	5	AL929469	AL929469 Zebrafish
c 38	70.6	4.7	212722	2	AC098909	AC098909 Rattus no
c 39	70.2	4.7	2167	3	PRSTARPA	Z30339 P. reichenow
c 40	69.6	4.6	202182	2	AC120639	AC120639 Rattus no
c 41	69.6	4.6	274060	2	AC095184	AC095184 Rattus no
c 42	69.4	4.6	203271	2	BX322566	BX322566 Danio rer
c 43	69.2	4.6	11534	3	PFARP1PR	Y08926 P. falciparu
c 44	69	4.6	178664	2	BX511115	BX511115 Danio rer
c 45	69	4.6	180023	2	BX897685	BX897685 Danio rer
c 46	68.8	4.6	333321	3	AC116986	AC116986 Dictyoste
c 47	68.6	4.5	209387	2	AC096448	AC096448 Rattus no
c 48	68.2	4.5	14626	3	EMOFLCEG	M76430 Bombyx mori
c 49	68	4.5	253151	3	AE014842	AE014842 Plasmodiu
c 50	68	4.5	293650	1	AP004172	AP004172 Mycoplasma
c 51	67.8	4.5	146570	3	AC117072	AC117072 Dictyoste
c 52	67.8	4.5	234117	2	AC130985	AC130985 Rattus no
c 53	67.6	4.5	169049	2	BX322580	BX322580 Danio rer
c 54	67.2	4.5	206779	2	BX511037	BX511037 Danio rer
c 55	66.8	4.4	115758	9	AC104634	AC104634 Homo sapi
c 56	66.8	4.4	165118	2	AC119053	AC119053 Danio rer
c 57	66.2	4.4	125623	3	AC115599	AC115599 Dictyoste
c 58	66	4.4	187846	2	AC146995	AC146995 Arabacia p
c 59	65.8	4.4	185107	2	AC118513	AC118513 Rattus no
c 60	65.8	4.4	243570	2	AC125757	AC125757 Rattus no
c 61	65.6	4.3	110000	2	PFMAL13_24	Continuation (25 o
c 62	65.4	4.3	186935	2	AC022322	AC022322 Homo sapi
c 63	65.2	4.3	114226	2	AC004710	AC004710 Plasmodiu
c 64	65.2	4.3	196490	2	AC005507	AC005507 Plasmodiu
c 65	65.2	4.3	210950	10	AC129317	AC129317 Mus muscu



QY	901	AAGGGAATAAATACATCACTACAGTCACTTATACAGAGATATTAAAGAGTGGAGCTAAACCA	960
Db	901		
QY	961	GGTTCAGCTGATTTACAGAAATCAACATTTGCGACCATCAACCCCAATACTAGCAAT	1020
Db	961		
QY	1021	GATGACCCAGGTCAAAAGTAAACAGTGGAGGTGTCACAAATTAATCTATATAAAAAAATTGAT	1080
Db	1021		
QY	1081	GGTTCACAAAAGCTTCATTACAGGTGCTATATTGTTTAAAGAACTGCTACGGGTCAA	1140
Db	1081		
QY	1141	TTTCTAACTTTAAACGATCAAAATACGTTGAATGGGCGACAGAACTAATGCAACGAA	1200
Db	1141		
QY	1201	TATACACAGGAGCAGATGGTATAATTACATTACAGGCTTGAAAGAGGTACATCTAT	1260
Db	1201		
QY	1261	CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATACTCTCAGAGGTTATT	1320
Db	1261		
QY	1321	TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTAGTTAAACCACTGTT	1380
Db	1321		
QY	1381	GAAAATAACAAAGTACTGAGTTGCTTCAACAGGTGGTATTGGTACAACAATTTCTAC	1440
Db	1381		
QY	1441	ATTATAGTGCATTTTATAGTAATAGGACAGGTATCGTGGCTTGGCTCGTCTGTTTA	1500
Db	1441		
QY	1501	CGTCTCTAA	1509
Db	1501		
RESULT 2	AX476887	Sequence 15 from Patent W0212294.	PAT 12-AUG-2002
LOCUS	AX476887	1509 bp	DNA
DEFINITION	AX476887	Sequence 15 from Patent W0212294.	
ACCESSION	AX476887		
VERSION	AX476887.1	GI:22216151	
KEYWORDS	Streptococcus agalactiae		
SOURCE	Streptococcus agalactiae		
ORGANISM	Streptococcus agalactiae		
REFERENCE	1	Streptococcus.	
AUTHORS	1	Aderson,E. and Bohnsack,J.	
TITLE	Group b streptococcus polypeptides nucleic acids and therapeutic compositions and vaccines thereof		
JOURNAL	Patent: WO 0212294-A 15 14-FEB-2002;		
ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US) ; University of Utah Research Foundation (US)			
FEATURES	Location/Qualifiers		
source	1..1509		
	/organism="Streptococcus agalactiae"		
	/mol_type="unassigned DNA"		
CDS	1..1509		
	/db_xref="taxon:1311"		
	/note="unnamed protein product"		
	/codon_start=1		
	/transl_table=11		

/protein_id="CAD43673.1"	
/db_xref="GI:22216152"	
/translation="MKKKMIOSLLVASLAFGMVSPVTPPIAFAAETGTTITVQDTOKGA	
TYKAYKVFDAIDNANVSDNKGASVLIPOKEAEYKASDENSLEFTTTTNGGRTYV	
TKKDTASANEIATWAKSISANTTPVSTVTESSNDGTENVISQYGYVSVTVANGAV	
IMVTSVTPNATIEHKNTDASATWDGGGKVDQKTSYVGDVVKITTYKNAVHTGERV	
YQYVTKDTPSASVVDLNEGSYEVTITDGSNITLLTQSGEKATTKYNLLBENNNTI	
TIPWAAVNTPTGNQANDDFFYKINTITVYTLVGLKSGKPSADLPENTNIATI	
NPNTSDPQKQVTRDQKITIKKIDGSTRASLOQAFVLKNAQGFNFNDTNVNEW	
GTEANATYTTGADGIIITGLKEGTVLVVEKAPLGVNLLDNQKVLILGDAVDTTN	
SDNLLVNPVTENNKGTELPSGGIGTTFYFIIGAILVIGAGIVLVARRRLRS"	
ORIGIN	
Query Match	100.0%; Score 1509; DB 6; Length 1509;
Best Local Similarity	100.0%; Pred. No. 3e-238;
Matches 1509; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATGAAAAGAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60
Db	
QY	1 ATGAAAAGAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60
Db	
QY	61 TCACCAAGTTAGCCGATAGCTTTTGGCGCTCAGACAGGACAAATTACAGTTCAAGATACT 120
Db	
QY	121 CAAAAAGCGCAACTATAAAGCATATAAAGTTTTTGTATGCAGAAATAGATAATGCAAA 180
Db	
QY	181 GTATCTGATTCGAATAAGATGGAGCTTCTTATTATTCCTCAAGGTAAAGAGCTGAG 240
Db	
QY	241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACCACTACTTAATGAGGGGAGA 300
Db	
QY	301 TATGTAACATAAAAAGATPACTGCGTCAGCAATGAGATTGCGACATGGGCTAAATCTATA 360
Db	
QY	361 TCAGCTAATACTACACCGTTTCCACTGTTACTGAGTCAAAATATGATGGTACTGAGTT 420
Db	
QY	421 ATTAATGTTTCCCAATATGGATATTATTATGTTTCTAGCACTGTTTAATAATGGAGCTGTA 480
Db	
QY	481 ATTATGGTTACATCTGTAACCTCCAAATGCTACTATTATCATGAAAAGAAATPACTGATCGGACA 540
Db	
QY	541 TGGGAGATGGTGGGAAAAACCTGTAGATCAAAAACGTTACTCGTGGTGGTATACATC 600
Db	
QY	601 AAATATCTATTACTTATAAGAAATGCAGTCAATTTATCATGTAAGAAAAGTGTATCAA 660
Db	
QY	661 TATGTTTAAAGGATACTATGCCATCTGCTTCTGTAGTTGATTGAAAGAGGCTCTTAT 720
Db	
QY	721 GAAGTAACTATTACTGATGGATCAGGGAATATTACAATCTCTAACTCAAGGTTTCGAAAAA 780
Db	
QY	781 GCAACTGGGAAGTATAACCTGTTAGAGGAAATATAATTTTCAGATTACTATTCCTGGTGG 840
Db	

```
QY 841 GCAGCTACCAATCTCCACCGGAATACTCAAAATGGAGCTAAATGATGACTTTTTTAT 900
Db 841 GCAGCTACCAATCTCCACCGGAATACTCAAAATGGAGCTAAATGATGACTTTTTTAT 900
QY 901 AAGGGAATAAATAACAATCAGTCAGTCCTTATACAGGAGTATTAAGAAGTGGAGCTAAACCA 960
Db 901 AAGGGAATAAATAACAATCAGTCAGTCCTTATACAGGAGTATTAAGAAGTGGAGCTAAACCA 960
QY 961 GGTTCAGCTGATTTACAGAGAAATACAAACATTCGACCATCAACCCCAATAGTCAAT 1020
Db 961 GGTTCAGCTGATTTACAGAGAAATACAAACATTCGACCATCAACCCCAATAGTCAAT 1020
QY 1021 GATGACCCAGGTCACCAAGTAACAGTCAGGAGTGGTCAAAATCTATAAAAAAATTTAT 1080
Db 1021 GATGACCCAGGTCACCAAGTAACAGTCAGGAGTGGTCAAAATCTATAAAAAAATTTAT 1080
QY 1081 GGTTCACAAAAGCTTCATTACAGGTCGTATATTTGTTTTAAAGAAATCTACGGGTCAA 1140
Db 1081 GGTTCACAAAAGCTTCATTACAGGTCGTATATTTGTTTTAAAGAAATCTACGGGTCAA 1140
QY 1141 TTTCTAACTTTAACGATACAAATACGTTGAATGGGCACAGAGCTAATGCAACAGAA 1200
Db 1141 TTTCTAACTTTAACGATACAAATACGTTGAATGGGCACAGAGCTAATGCAACAGAA 1200
QY 1201 TATACACAGGAGCAGATGGTAAATACCATACAGGCTTGAAGAAGGTACATACAT 1260
Db 1201 TATACACAGGAGCAGATGGTAAATACCATACAGGCTTGAAGAAGGTACATACAT 1260
QY 1261 CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAACTCTCAGAAGGTTAT 1320
Db 1261 CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAACTCTCAGAAGGTTAT 1320
QY 1321 TTAGGAGTGGAGCCACTGATACGACTTAATTCAGATAACCTTTAGTTAACCAACTGTT 1380
Db 1321 TTAGGAGTGGAGCCACTGATACGACTTAATTCAGATAACCTTTAGTTAACCAACTGTT 1380
QY 1381 GAAAATAACAAAGGTAAGTGGTTCCTTCAACAGGTCGTATTCGTACAAACAATTTCTAC 1440
Db 1381 GAAAATAACAAAGGTAAGTGGTTCCTTCAACAGGTCGTATTCGTACAAACAATTTCTAC 1440
QY 1441 ATTATAGTGCAATTTTAGTAATAGAGCAGAGTATCGTGTGTCGTCGTCGTTTA 1500
Db 1441 ATTATAGTGCAATTTTAGTAATAGAGCAGAGTATCGTGTGTCGTCGTCGTTTA 1500
QY 1501 CGTTCCTAA 1509
Db 1501 CGTTCCTAA 1509

RESULT 3
AC116998
LOCUS AC116998 216959 bp DNA linear ROD 08-NOV-2003
DEFINITION Mus musculus BAC clone RP23-16F6 from 18, complete sequence.
ACCESSION AC116998
VERSION AC116998.3 GI:24080765
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 216959)
AUTHORS Ieak,A., Kozlowski,A., Spalding,L. and Mangiapanello,L.
TITLE The sequence of Mus musculus BAC clone RP23-16F6
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 216959)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 216959)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2002) Genome Sequencing Center, 4444 Forest Park
```

Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 216959)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 216959)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (17-OCT-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
6 (bases 1 to 216959)  
Wilson,R.  
Direct Submission  
Submitted (08-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 17, 2002 this sequence version replaced gi:21218578.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: M\_BA0016F06

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RP23-23 BAC Library has been constructed by Kazutoyo Oseawa and Minako Tateo in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

#### FEATURES

source

1. 216959  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="18"  
/map="18"  
/clone="RP23-16F6"  
/clone\_lib="RP23-23"  
1271..1406  
/rpt\_family="ERVK"  
1423..1535  
/rpt\_family="B4"  
1611..1806  
/rpt\_family="RMERIA"  
3133..3210  
/rpt\_family="ERV1"  
3256..3331  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region













zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zK46N5  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 111658 bases at least Q40  
Consensus quality: 112600 bases at least Q30  
Consensus quality: 113615 bases at least Q20  
Insert size: 115507; sum-of-contigs  
Insert size: 12756; 0.3% error; agarose-fp  
Quality coverage: 7.13x in Q20 bases; sum-of-contigs Quality  
coverage: 6.91x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2116: contig of 2116 bp in length  
\* 2117 2216: gap of 100 bp  
\* 2217 8578: contig of 6362 bp in length  
\* 8579 8678: gap of 100 bp  
\* 8679 14990: contig of 6312 bp in length  
\* 14991 26940: gap of 100 bp  
\* 26941 27041: contig of 11850 bp in length  
\* 27042 34475: contig of 7435 bp in length  
\* 34476 34575: gap of 100 bp  
\* 34576 41111: contig of 6536 bp in length  
\* 41112 41211: gap of 100 bp  
\* 41212 52851: contig of 11640 bp in length  
\* 52852 52951: gap of 100 bp  
\* 52952 58237: contig of 5286 bp in length  
\* 58238 58337: gap of 100 bp  
\* 58338 73639: contig of 15302 bp in length  
\* 73640 73739: gap of 100 bp  
\* 73740 96363: contig of 22624 bp in length  
\* 96364 96463: gap of 100 bp  
\* 96464 99575: contig of 3112 bp in length  
\* 99576 104567: gap of 100 bp  
\* 104568 104667: contig of 4892 bp in length  
\* 104668 109346: contig of 100 bp  
\* 109347 109446: contig of 4679 bp in length  
\* 109447 116807: contig of 100 bp  
\* 116808 116807: contig of 7361 bp in length.  
Location/Qualifiers  
1. .116807  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone\_lib="DKEY-46N5"  
/clone\_lib="DanioKey"  
1. .2116  
/notes="assembly fragment:00043  
fragment\_chain:1"  
misc\_feature  
2217. .8578  
/note="assembly fragment:00104  
fragment\_chain:1"  
misc\_feature  
8679. .14990  
/note="assembly fragment:00322  
fragment\_chain:1"  
misc\_feature  
15091. .76940  
/note="assembly fragment:00769  
fragment\_chain:1"

misc\_feature 27041. .34475  
/note="assembly fragment:00537  
fragment\_chain:1"  
misc\_feature 34576. .41111  
/note="assembly fragment:00392  
fragment\_chain:1"  
misc\_feature 41212. .52851  
/note="assembly fragment:00633  
fragment\_chain:2"  
misc\_feature 52952. .58237  
/note="assembly fragment:00201  
fragment\_chain:2"  
misc\_feature 58338. .73639  
/note="assembly fragment:00925  
fragment\_chain:2"  
misc\_feature 73740. .96363  
/note="assembly fragment:01125  
fragment\_chain:2"  
misc\_feature 96464. .99575  
/note="assembly fragment:00066.0"  
misc\_feature 99676. .104567  
/note="assembly fragment:00151"  
misc\_feature 104668. .109346  
/note="assembly fragment:00258"  
misc\_feature 109447. .116807  
/note="assembly fragment:00464"  
ORIGIN  
Query Match 5.7%; Score 86; DB 2; Length 116807;  
Best Local Similarity 44.7%; Pred.No.3.1e-05;  
Matches 335; Conservative 0; Mismatches 415; Indels 0; Gaps 0;  
QY 280 ACTACTAATGGAGGAGCAATATGTAATACTAAAGAAAGTACGCGTCAGCAAAATGAGATT 339  
Db 26253 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 26194  
QY 340 GGCACATGGGCTAAATCTATATCAGCTAATACACCACTGTTCCAGTGTCTACTGAGTCA 399  
Db 26193 GATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 26134  
QY 400 AATAATGATGGTACTGAGGTTAATGTTTCCCAATATGATATATATTTCTTCTAGC 459  
Db 26133 AATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 26074  
QY 460 ACTGTTAATAATGGAGCTGTATATATGTTTCACTCTGTAACCTCAAACTCTACTATTCAT 519  
Db 26073 AATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 26014  
QY 520 GAAAAGATCTGATCGGACATGGGAGATGGTGGGAAAGAACTGTAGATCAAAAACG 579  
Db 26013 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 25954  
QY 580 TACTCGTTGGTGATACAGTCAAAATATCTATTACTTAAAGAAATGCAAGTCAATATCAT 639  
Db 25953 AATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 25894  
QY 640 GGTACAGAAAAAGTGTATCAATATGTTTAAAGGATACCTATGCCATCTCTCTCTGAGTT 699  
Db 25893 AATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 25834  
QY 700 GATTTGAACGAGGGCTTATGAAGTACTATTACTGATGGATCAGGGAATATTACAAT 759  
Db 25833 AATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 25774  
QY 760 CTAACCTCAAGGTTCCGAAAAAGCACTGGGAAAGTATAACCTGTGAGGAAATAATAAT 819  
Db 25773 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 25714  
QY 820 TTCACGATTACTATTCGTTGGGAGCTACCAATATCTTCAACCGGAAATCTCAAAATGGA 879  
Db 25713 GATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 25654  
QY 880 GCTAATGATGACTTTTTTTTATTAAGGAATAATAACAATCACAGTCACTTATACAGGAGTA 939



```

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
Wray, P.
Direct Submission
Submitted (10-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 12, 2003 this sequence version replaced gi:37606167.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zK245C12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 19717 bases at least Q40
Consensus quality: 198152 bases at least Q30
Consensus quality: 198296 bases at least Q20
Insert size: 198443; sum-of-contigs
Insert size: 211631; 3.9% error; agarose-fp
Quality coverage: 8.87x in Q20 bases; sum-of-contigs Quality
coverage: 9.18x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 55329: contig of 55329 bp in length
* 55330 55429: gap of 100 bp
* 190714: contig of 135285 bp in length
* 190715 190814: gap of 100 bp
* 190815 196496: contig of 5682 bp in length
* 196497 196596: gap of 100 bp
* 196597 198743: contig of 2147 bp in length.
*
* Location/Qualifiers
* 1..198743
* /organism="Danio rerio"
* /mol_type="genomic DNA"
* /db_xref="taxon:7955"
* /clone_DKEY="245C12"
* /clone_lib="DanioKey"
*
* 1..55329
* /note="assembly fragment:03483"
* fragment_chain:1
* 55430..190714
* /note="assembly fragment:01158"
* fragment_chain:1
* 190815..196496
* /note="assembly fragment:02910"
* 196597..198743
* /note="assembly fragment:03681"
*
* ORIGIN
*
* Query Match 5.6%; Score 84; DB 2; Length 198743;
* Best Local Similarity 45.5%; Pred. No. 5.5e-05;
* Matches 300; Conservative 0; Mismatches 360; Indels 0; Gaps 0;
*
* QY 243 TAAGCTTCACCTGATTTTATCTCTTTTACGACAACTACTAATGGAGGAGACATA 302
* DB 86200 TACTACTACTCTTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86141
*
* QY 303 TGTTACTAAAAAGATCTGGCTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATC 362
* DB 86140 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86081

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

FEATURES  
source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

ORIGIN

Query Match 5.6%; Score 84; DB 2; Length 198743;  
Best Local Similarity 45.5%; Pred. No. 5.5e-05;  
Matches 300; Conservative 0; Mismatches 360; Indels 0; Gaps 0;

QY 243 TAAGCTTCACCTGATTTTATCTCTTTTACGACAACTACTAATGGAGGAGACATA 302  
DB 86200 TACTACTACTCTTATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86141

QY 303 TGTTACTAAAAAGATCTGGCTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATC 362  
DB 86140 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86081

QY 363 AGCTAATACTACACCAGTTTCCACTGTTACTGACTCAAAATAATGATGGTACTGAGGTTAT 422  
DB 86080 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86021

QY 423 TAATGTTTCCCAATATGATATTAATGTTTCTAGCACTGTTTAATATGAGGCTGTAAT 482  
DB 86020 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85961

QY 483 TATGTTTACATCTGTAACTCCAAATGCTACTATTTCATGAAAGAATACTGATGCGACATG 542  
DB 85960 TAATGATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85901

QY 543 GGGAGATCGTGTGGAAAACTGTAGATCAAAAACGCTACCTCGGTTGGTGTATACAGTCAA 602  
DB 85900 TAAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85841

QY 603 ATATACTATTCTTATAGAACTGAGTCAATATCATGTTACAGAAAGTGTATCAATA 662  
DB 85840 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85781

QY 663 TGTATAAAGGATCACTATGCCATCTGCTTCTGTAGTTGATTGAAACGAAGGCTTTATGA 722  
DB 85780 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85721

QY 723 AGTAATACTATTGATGAGTCAAGGAATATTACAACTCTCACTCAAGGTCGGAAGAGC 782  
DB 85720 TAAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85661

QY 783 RACTGGGAAGTATAACCTGTTAGAGAAAAATAATAATTTCCAGTACTATTCCGTGGGC 842  
DB 85660 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85601

QY 843 AGCTACCACTACTCCAAACCGGAATACTCAAAATGGAGTAAATGATGATCTTTTATATA 902  
DB 85600 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85541

RESULT 10  
AC007926/c

LOCUS

DEFINITION

ACCESSION

VERSION

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Location/Qualifiers

1..198743

/organism="Danio rerio"

/mol\_type="genomic DNA"

/db\_xref="taxon:7955"

/clone\_DKEY="245C12"

/clone\_lib="DanioKey"

1..55329

/note="assembly fragment:03483"

fragment\_chain:1

55430..190714

/note="assembly fragment:01158"

fragment\_chain:1

190815..196496

/note="assembly fragment:02910"

196597..198743

/note="assembly fragment:03681"

ORIGIN

Query Match 5.6%; Score 84; DB 2; Length 198743;  
Best Local Similarity 45.5%; Pred. No. 5.5e-05;  
Matches 300; Conservative 0; Mismatches 360; Indels 0; Gaps 0;

QY 243 TAAGCTTCACCTGATTTTATCTCTTTTACGACAACTACTAATGGAGGAGACATA 302  
DB 86200 TACTACTACTCTTATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86141

QY 303 TGTTACTAAAAAGATCTGGCTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATC 362  
DB 86140 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86081



source

1. .155204

/organism="Trypanosoma brucei"

/mol\_type="genomic DNA"

/isolate="GUTat10.1"

/db\_xref="taxon:5691"

/chromosome="11"

/clone="RPC193-3H15"

ORIGIN

Query Match

Best Local Similarity

5.5%; Score 83.4; DB 2; Length 155204;

44.6%; Pred. No. 7.5e-05;

Matches 327; Conservative

0; Mismatches 406; Indels

0; Gaps

0;

Qy

170 ATAATGCAATGTATCTGATTCGATAAAGATGGAGCTTCATTATTAATTCCTCAAGGTA 229

Db

81317 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81258

Qy

230 AAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATCTCTTTTACGACAACACTACTAATG 289

Db

81257 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81198

Qy

290 GAGGAGACAATATGTAACATAAAGATACCTGCGTCAGCAATGAGATTGCGACATGGG 349

Db

81197 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81138

Qy

350 CTAATCTATATCAGCTAATACACACCGTTCCACTGTTACTGAGTCAATAATGATG 409

Db

81137 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81078

Qy

410 GTACTGAGCTTAAATGTTTCCCAATATGGATATTTATTTCTAGCACTGTTAATA 469

Db

81077 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81018

Qy

470 AFGGAGCTGTATATGTTTACATCTGTAACCTCCAAATGCTACTATTCATGAAAGAATA 529

Db

81017 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80958

Qy

530 CTGATCGACATGGGGAGATGGTGGAAAACTGTAGATCAAAAACGGTACTCGGTTG 589

Db

80957 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80898

Qy

590 GTGATACAGTCAAAATATCTATTACTTATAAGATGCGATCAATATCATGTCAGAGAA 649

Db

80897 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80838

Qy

650 AAGTGTATCAATATGTTATAAAGGATCTATGCCATCTGCTTCTGTAGTTGATTTGAACG 709

Db

80837 AAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80778

Qy

710 AAGGTCTTATGAAGTAACTATTACTGATGATCAGGGAATATTCAACTCTAAGCTCAAG 769

Db

80777 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80718

Qy

770 GTTCGGAAAAACCACTGGAGTATACCTGTTAGAGGAAATAATAATTTTCAGATTA 829

Db

80717 ATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80658

Qy

830 CTATTCCGTGGCAGCTACCAATCTCCAAACCGGAAATACTCAAAATGGAGCTAATGATG 889

Db

80657 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80598

Qy

890 ACTTTTTTTATAA 902

Db

80597 ATGATAATAATAAA 80585

RESULT 11

AL929250/c

LOCUS

94534 bp DNA linear VRT 16-SEP-2003

DEFINITION

Zebrafish DNA sequence from clone CH211-137A2, complete sequence.

ACCESSION

AL929250

VERSION

AL929250.8 GI:34787249

KEYWORDS

HTG.

SOURCE

Danio rerio (zebrafish)

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 94534)

Almeida, J.

Direct Submission

Submitted (16-SEP-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Sep 16, 2003 this sequence version replaced gi:27848021.

-----

Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: zfish-help@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/CH211-137A2 is from a CHORI-211 BAC library VECTOR: PTARBAC2.1

Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D\_rerio/fishmask.shtml.

FEATURES

Location/Qualifiers

1. .94534

/organism="Danio rerio"

/mol\_type="genomic DNA"

/db\_xref="taxon:7955"

/clones="CH211-137A2"

/clone\_lib="CHORI-211"

ORIGIN

Query Match

Best Local Similarity

5.5%; Score 82.4; DB 5; Length 94534;

44.5%; Pred. No. 0.00013;

Matches 326; Conservative

0; Mismatches 406; Indels

0; Gaps

0;

Qy

103 ATTACAGTTCAGATACCTCAAAAAGCGCAACCTATAAGCATATAAGTTTGTGATCA 162

Db

30450 AATATAGTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 30391

Qy

163 GAATAGATTAATGCAATGATCTGATTCGATAAAGATGGAGCTTCTTATTATTCCT 222

Db

30390 AATAATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 30331

Qy

223 CAAGGTAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAACTCTCTTTTACGCAACT 282

Db

30330 AATAATAATAATGATGATAATAAAGATAATAATAATAATAATAATAATAATAATA 30271

283 ACTAATGGAGGAGAACATATGTAACTAAAAAGATGCTGCTCAGCAATGAGATTCCG 342  
30270 AATAATGATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 30211  
343 ACATGGGCTAATCTATATCAGCTAATCTACACAGGTTTCCACCTGTTACTGAGTCAAT 402  
30210 AATAATAGTAATAATAATGATAATAATGATAATAATAATAATAATAATAATAATAAT 30151  
403 AATGATGGTACTGAGGTTAATTAATGTTTCCCAATATGGAATATTAATGTTTCTAGCACT 462  
30150 AATAATGGTGATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAAC 30091  
463 GTTAATAATGGAGCTGTAATATGTTTACATCTGTAATCTCCAAATGCTACTATTTCATGAA 522  
30090 AGTAATAATAATGATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAGT 30031  
523 AAGAATACTGATCGATCGGAGATGGTGGGAAAACTGTAGATCAAAAAACGTAC 582  
30030 AATAATAATGATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAAT 29971  
583 TCGTGGTGATACAGTCAATATATCTATTACTTTAAGAAATGCAATGATCATGCT 642  
29970 AATAATAATGGTGATGATAATAATAATAATAATAATAATAATAATAATAATAATGGT 29911  
643 ACAGAAAAAGTATCAATATGTTTATAAAGGATGATGCTGCTCTCTGTTAGTTGAT 702  
29910 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 29851  
703 TTGAACGAAGGCTTATGAACTAATCTATTACTGATGATGATGATGATGATGATGATGAT 762  
29850 AGTAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAGCGAT 29791  
763 ACTCAAGTTCCGAAAAAGCACTGGGAGTATACCTGTTAGAGGAAATAATAATTTTC 822  
29790 AATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGAT 29731  
823 ACGATTACTATT 834  
29730 AATAATTAAT 29719

RESULT 12

PFMAL13\_07

WPCOMMENT

Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509

Fragment Name	Begin	End
PFMAL13_00	1	110000
PFMAL13_01	100001	210000
PFMAL13_02	200001	310000
PFMAL13_03	300001	410000
PFMAL13_04	400001	510000
PFMAL13_05	500001	610000
PFMAL13_06	600001	710000
PFMAL13_07	700001	810000
PFMAL13_08	800001	910000
PFMAL13_09	900001	1010000
PFMAL13_10	11000001	1110000
PFMAL13_11	11000001	1210000
PFMAL13_12	12000001	1310000
PFMAL13_13	13000001	1410000
PFMAL13_14	14000001	1510000
PFMAL13_15	15000001	1610000
PFMAL13_16	16000001	1710000
PFMAL13_17	17000001	1810000
PFMAL13_18	18000001	1910000
PFMAL13_19	19000001	2010000
PFMAL13_20	20000001	2110000
PFMAL13_21	21000001	2210000
PFMAL13_22	22000001	2310000
PFMAL13_23	23000001	2410000
PFMAL13_24	24000001	2510000
PFMAL13_25	25000001	2610000
PFMAL13_26	26000001	2710000

PFMAL13\_27 2732359  
Continuation\_ (8 of 28) of PFMAL13 from base 700001 (AL844509 Plasmodium falciparum 3D7 cl

Query Match 5.3%; Score 80.6; DB 2; Length 110000;  
Best Local Similarity 44.8%; Pred. No. 0.00024;  
Matches 308; Conservative 0; Mismatches 379; Indels 0; Gaps 0;

QY 137 ATAAAGCATATAAAGTTTGGATGCAGAAATAGATAATGCAATGTATCTGATTCGAATA 196  
Db 43055 ATGATGATGATAATAATAATGATGATGATGATAATAATAATGATGATGATAATAATA 43114  
QY 197 AAGATGAGCTTCTATTATTTTCTCAAGGTAAAGAGCTGAGTATAAGCTTCAACTG 256  
Db 43115 ATGATGATGATAATAATAATGATGATGATGATAATAATAATGATGATGATGATAATA 43174  
QY 257 ATTTTATTTCTCTTTTACGCAACTACTAATGGAGGAGAACATATGTAACATAAAAAAG 316  
Db 43175 ATAATGATGATGATGATAATAATAATGATGATGATGATAATAATAATGATGATGATA 43234  
QY 317 ATACTGGCTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAGCTAACTACAC 376  
Db 43235 ATAATAATGATGATGATAATAATAATGATGATGATAATAATAATGATGATGATG 43294  
QY 377 CAGTTTCCACTGTTACTGAGTCAAAATGATGATGATGATGATGATGATGATGATGATG 436  
Db 43295 ATAATAATGATGATGATAATAATAATGATGATGATGATGATGATGATGATGATGATG 43354  
QY 437 ATGGATATTATTATGTTTCTAGCCTGTTAATTAATGGAGCTGTAATTTATGGTTACATCTG 496  
Db 43355 ATGATAATAATAATGATGATGATAATAATAATGATGATGATAATAATAATGATGATG 43414  
QY 497 TAACTCCAAATGCTACTATTTCATGAAAGATATCTGATGCGACATGGGAGATGGTGTG 556  
Db 43415 ATGATGATAATAATAATGATGATGATAATAATAATGATGATGATGATGATGATGATG 43474  
QY 557 GAAAACTGTAGATCAAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATATCTATTACT 616  
Db 43475 ATGATGATGATAATAATAATGATGATGATAATAATAATGATGATGATAATAATAATA 43534  
QY 617 ATAGAATGCACTCAATTTATCATGGTACAGAAAAGTGATCAATNTGTTATAAGGATA 676  
Db 43535 ATGATGATGATGATAATAATAATGATGATGATAATAATAATGATGATAATAATAATCAT 43594  
QY 677 CTATGCCATCTGCTTCTGTTAGTTGATTTGAACGAGGCTTCTTGAAGTAATCTATTACTG 736  
Db 43595 ATAAATCATATAATAATAATAATTTATTAATAATACATAATAATAATAATAATAATAAG 43654  
QY 737 ATGGATCAGGGAATATTACAACTCTAACTCAAGGTTCCGAAAAAGCAACTGGAGATATA 796  
Db 43655 AAGAAGTATCATTTAATATATCTGAACCCCCCAATGCTATAGGAATAATAATGGTAATT 43714  
QY 797 ACTGTTAGAGGAAATAATAATAATTTTCA 823  
Db 43715 ATATTTCCAGGATGAAGATAATAATTACA 43741

RESULT 13

PFMAL13\_07

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

225581 bp DNA linear HTG 06-AUG-2003  
BX537105  
Danio rerio clone DKEY-3L24, 18 unordered pieces.  
BX537105  
BX537105.2 GI:31559379  
HTG; HTGS PHASE1; HTGS CANCELLED.  
Danio rerio (zebrafish)  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 225581)  
Burton, J.  
Direct Submission  
Submitted (05-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jun 9, 2003 this sequence version replaced gi:31441851.

COMMENT

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zK3L24  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 217427 bases at least Q40  
Consensus quality: 219130 bases at least Q30  
Consensus quality: 220818 bases at least Q20  
Insert size: 223881; sum-of-contigs  
Insert size: 219000; 7.3% error; agarose-fp  
Quality coverage: 6.82x in Q20 bases; sum-of-contigs Quality  
coverage: 7.43x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 20602: contig of 20602 bp in length  
\* 20603 20702: gap of 100 bp  
\* 20703 57831: contig of 37129 bp in length  
\* 57832 57931: gap of 100 bp  
\* 57932 65164: contig of 7233 bp in length  
\* 65165 65264: gap of 100 bp  
\* 65265 71392: contig of 6128 bp in length  
\* 71393 71492: gap of 100 bp  
\* 71493 93677: contig of 22185 bp in length  
\* 93678 93777: gap of 100 bp  
\* 93778 101598: contig of 7821 bp in length  
\* 101599 128689: gap of 100 bp  
\* 128690 128789: contig of 26991 bp in length  
\* 128790 138632: contig of 9843 bp in length  
\* 138633 138732: gap of 100 bp  
\* 138733 144193: contig of 5461 bp in length  
\* 144194 144293: gap of 100 bp  
\* 144294 155818: contig of 11525 bp in length  
\* 155819 155918: gap of 100 bp  
\* 155919 168354: contig of 12436 bp in length  
\* 168355 168454: gap of 100 bp  
\* 168455 171301: contig of 2847 bp in length  
\* 171302 171401: gap of 100 bp  
\* 171402 180364: contig of 8963 bp in length  
\* 180365 180464: gap of 100 bp  
\* 180465 188604: contig of 8140 bp in length  
\* 188605 188705: gap of 100 bp  
\* 188705 192855: contig of 4151 bp in length  
\* 192856 192955: gap of 100 bp  
\* 192956 201754: contig of 8799 bp in length  
\* 201755 201854: gap of 100 bp  
\* 201855 211372: contig of 9518 bp in length  
\* 211373 211472: gap of 100 bp  
\* 211473 225581: contig of 14109 bp in length.

FEATURES

Source

1..225581  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-3L24"  
/clone\_lib="DanioKey"  
1..20602  
/note="assembly fragment:01463  
fragment\_chain:1"

misc\_feature

misc\_feature 20703..57831  
/note="assembly fragment:01892  
fragment\_chain:1"  
misc\_feature 57932..65164  
/note="assembly fragment:02319  
fragment\_chain:1"  
misc\_feature 65285..71392  
/note="assembly fragment:02166  
fragment\_chain:1"  
misc\_feature 71493..93677  
/note="assembly fragment:01041  
fragment\_chain:1"  
misc\_feature 93778..101598  
/note="assembly fragment:02235  
fragment\_chain:1"  
misc\_feature 101699..128689  
/note="assembly fragment:01236  
fragment\_chain:1"  
misc\_feature 128790..138632  
/note="assembly fragment:02758  
fragment\_chain:1"  
misc\_feature 138733..144193  
/note="assembly fragment:02245  
fragment\_chain:1"  
misc\_feature 144294..155818  
/note="assembly fragment:01852  
fragment\_chain:1"  
misc\_feature 155919..168354  
/note="assembly fragment:01722  
fragment\_chain:1"  
misc\_feature 168455..171301  
/note="assembly fragment:00772  
fragment\_chain:2"  
misc\_feature 171402..180364  
/note="assembly fragment:02453  
fragment\_chain:2"  
misc\_feature 180465..188604  
/note="assembly fragment:00524  
fragment\_chain:2"  
misc\_feature 188705..192855  
/note="assembly fragment:02841  
fragment\_chain:2"  
misc\_feature 192956..201754  
/note="assembly fragment:00644  
fragment\_chain:2"  
misc\_feature 201855..211372  
/note="assembly fragment:02536  
fragment\_chain:2"  
misc\_feature 211473..225581  
/note="assembly fragment:01373  
fragment\_chain:2"

ORIGIN

Query Match 5.3%; Score 80.4; DB 2; Length 225581;  
Best Local Similarity 45.3%; Pred. No. 0.00021;  
Matches 291; Conservative 0; Mismatches 351; Indels 0; Gaps 0;  
Qy 261 TAAATCTCTTTTACGACAACTACTAATGAGGGGAGACATATGTAACATAAGATAC 320  
Db 151722 TACTTCTACTACTACTACTACTTATAATAATAATAATAATAATAATAATA 151781  
Qy 321 TGGGTACGAAATGAGATTGGACATGGGCTAAATCTATATCAGTAACTACTACACAGT 380  
Db 151782 TAAATAATAATAATGATATAATAATAATAATAATAATAATAATAATAATAATA 151841  
Qy 381 TTCACTGTTACTAGTCAATAATAATGAGTACTGAGGTTTATTAATGTTTCCCATATGG 440  
Db 151842 TAAATAATAATAATGATATAATAATAATAATAATAATAATAATAATAATAATA 151901  
Qy 441 ATATTATTATGTTCTAGCACTGTTAAATAATGAGGCTGTAATTAATGTTTACATCTGTAAC 500  
Db 151902 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 151961

```
QY 501 TCCAAATGCTACTATTTCATGAAAGAATACTGATCGGACATGGGAGATGGTGGTGA 560
Db 151962 TATTGATAATAATGATAATAATAATAATGATAATAATAATAATAATAATAATA 152021
QY 561 AACTGTAGATCAAAAACCTACTCGGTGGTGATACAGTCAGTCAATATATCTATTACTTTATA 620
Db 152022 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 152081
QY 621 GAATGAGTCAATTTATCATGGTACAGAAAAGTGTATCAATATGTTTATAAGGATACAT 680
Db 152082 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 152141
QY 681 GCCATCTGTTCTGTAGTGAATTTGAACGAGGCTTATGAAGTAACTATTACTGATGG 740
Db 152142 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 152201
QY 741 ATCAGGGAATATTACAACTCTACTCAAGTTCGGGAAAAGCACTGGGAAAGTATAACCT 800
Db 152202 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 152261
QY 801 GTTAGAGGAAAATAATAATTCACGATTACTATTCGCGGGGAGCTACCAATCTCCAC 860
Db 152262 TAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 152321
QY 861 CGGAATACTCAAAATGGAGCTAATGATGACTTTTATATA 902
Db 152322 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 152363
```

```
RESULT 14
AF397467/c 10115 bp DNA linear VRT 27-NOV-2001
LOCUS Ictalurus punctatus novel immune-type receptor 2 (NITR2) and novel
DEFINITION immune-type receptor 4 (NITR4) genes, complete cds.
ACCESSION AF397467
VERSION AF397467.1 GI:17105096
KEYWORDS
SOURCE Ictalurus punctatus (channel catfish)
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
Ictaluridae; Ictalurus.
1 (bases 1 to 10115)
Hawke, N.A., Yoder, J.A., Haire, R.N., Mueller, M.G., Litman, R.T.,
Miracle, A.L., Stuge, T., Shen, L., Miller, N. and Litman, G.W.
Extraordinary variation in a diversified family of immune-type
receptor genes
Proc. Natl. Acad. Sci. U.S.A. 98 (24), 13832-13837 (2001)
JOURNAL MEDLINE 21574523
PUBMED 11698645
REFERENCE 2 (bases 1 to 10115)
Hawke, N.A., Yoder, J.A., Haire, R.N., Mueller, M.G., Litman, R.T.,
Miracle, A.L., Stuge, T., Miller, N. and Litman, G.W.
Direct Submission
Submitted (09-JUL-2001) Department of Pediatrics, University of
South Florida Children's Research Institute, 140 Seventh Avenue
South, St. Petersburg, FL 33701, USA
FEATURES
source
1..10115
/organism="Ictalurus punctatus"
/mol_type="genomic DNA"
/db_xref="taxon:7998"
<2836..>4401
/gene="NITR2"
join(<2836..2875,3025..3381,3759..4112,4226..>4401)
/gene="NITR2"
/product="novel immune-type receptor 2"
join(2836..2875,3025..3381,3759..4112,4226..4401)
/gene="NITR2"
/note="IpNITR2"
/codon_start=1
/product="novel immune-type receptor 2"
/protein_id="AAL35555.1"
```

```
/db_xref="GI:17105097"
/translation="MLTACFILLCLPLHLVKQTDFRNSDSLVAFAKEGDSVNISCIVE
SDMAFSPYKNLQCKPLISNFYKDKKATFHEFKNARFSWMNEKSKTNLEIKG
LQUSDATTFCSAHSNIIVEFGEGTFLVQASHLSVLQOPVHELAPHGSLVLTCT
VITDRGAGSHVYVFRHNSGRCSHPGVIYTHGDSNGRCKKNPKRAGSLTRMCVSLPKTN
LSTSDVGTGTHCAVACGQLFGNGTKIDMKITGLCVFENLIQTLVLFLFSVRSVLLCCV
ILMIYFCSSKKAKCSVRRTTKKNLAQLEK"
<7826..>9190
/gene="NITR4"
join(<7826..7865,8018..8368,8631..8975,9084..>9190)
/gene="NITR4"
/product="novel immune-type receptor 4"
join(7826..7865,8018..8368,8631..8975,9084..9190)
/gene="NITR4"
/note="IpNITR4"
/codon_start=1
/product="novel immune-type receptor 4"
/protein_id="AAL35555.1"
/db_xref="GI:17105098"
/translation="MILYFFLLCTSDVCLYIEEVMQEPLEITAEDEGGSVKLTCLYTK
QRTTIIWLROKLGEPYVAVTQVQPOAYYNI FEKNVFNARI GPSSFNLSINIK
LSDATYCAVTFLEYITFGQSTVLI VKDOSLKRTHVQOEKVMPEGDSVQCTIV
TDRGAGHNVHFRHSGSDSGIIVTQSGSGCKESFNASSSOTCVVSLPKNLRV
SDAGTYTCVAVVACGEILCGSEAKQNIKVHFOIEVLVLLSIIRSAVLLFIICLLCT
YIGH"
ORIGIN
Query Match 5.3%; Score 80; DB 5; Length 10115;
Best Local Similarity 47.5%; Pred. No. 0.00064;
Matches 270; Conservative 0; Mismatches 295; Indels 3; Gaps 1;
QY 171 TAATGCAAAATGATCTGATTCGAATAAAGATGGAGCTCTTATTTAATCTCTCAAGTAA 230
Db 6797 TACTACTACTATTACTGCTACTACTGCTGTTACTACTGCTGCTACTACTACTACTAC 6738
QY 231 AGAGCTGAGTATAAGCTTCAACTGATTTTAAATCTCTTTTACGACCAACTACTAATGG 290
Db 6737 TGCTACTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6678
QY 291 AGGGAGAACATATGTAATAAAGATGATCGGTGAGCAAAATGAGATGGGATGGGC 350
Db 6677 TACTAGTACTACTGCTACTACTATTACTGCTGCTACTACTACTACTACTACTACTAC 6618
QY 351 TAAATCTATATCAGTAATACTACACAGTTTCCACTGCTTACTGAGTCAAAATAATGATG- 409
Db 6617 TATTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTGT 6558
QY 410 --GTACTGAGGTATTAATATGTTTCCCAATATGGAATATTATGTTTCTAGACATGTTAA 467
Db 6557 TACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTGCTGCTGTTTAC 6498
QY 468 TAATGAGCTGTAATATGTTTACATCTGTAACCTCAATGCTACTATTTCATGAAAGAA 527
Db 6497 TACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTGTTAC 6438
QY 528 TACTGATCGACATGGGAGATGGTGGGAAAACTGTAGATCAAAAAACGCTACTCGGT 587
Db 6437 TGCTGCTACTACTAGTACTGCTGCTATTACTACTGCTGCTACTACTACTACTGCTACTAT 6378
QY 588 TGGTGATACAGTCAAAATATACTATTATTAAGAATGAGTCAATTATCATGTTGATGAA 647
Db 6377 GATTACTACTACTCTTCTGCTACTACTACTACTACTACTACTACTACTACTACTACTAC 6318
QY 648 AAAAGTGATCAATATGTTTATAAGGATACTATGCACTGCTGCTGCTGCTGCTGCTGCTG 707
Db 6317 TACTACTAATACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 6258
QY 708 CGAAGGCTCTTATGAAGTAACTATTACT 735
Db 6257 TACTACTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 6230
```



```
QY 305 TAACTAAAAGAGTACTGCGTCAGCAATGAGATTCCGACATGGGCTAAATCTATATCAG 364
D 167070 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 167011
QY 365 CTAATACTACACCAGTTTCCACTGTTTACTGAGTCAATATGATGCTAGTGGTTATTA 424
D 167010 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166951
QY 425 ATGTTTCCCAATATGATATTAATGTTTCTAGCACTGTTAATATGAGGCTGTAATTA 484
D 166950 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166891
QY 485 TGGTTACTCTGTAATCCAAATGCTACTATTTCATGAAAAGATACTGATCGCATGGG 544
D 166890 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166831
QY 545 GAGATGCTGTGGAATAAATCTGTAGATCAAAAACGTACTCGGTTGGTGATACAGTCAAT 604
D 166830 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166771
QY 605 ATACTATTACTTATAGATGCAAGTCAATATCATGCTACAGAAAAGTGTATCATATG 664
D 166770 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166711
QY 665 TTATAAGGATACTATGCCATCTGCTTCTGTAGTTGATTTGAACGAAGGCTCTTATGA 724
D 166710 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166651
QY 725 TAACTATTACTGATGATCAGGGAATATTACAATCTTAAGTCAAGGTTCCGAAAAGCAA 784
D 166650 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166591
QY 785 CTGGGAAGTAACTGTTAGAGAAAATTAATATTTCCAGTATTACTATTCCTGGGCGAG 844
D 166590 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166531
QY 845 CTACCAATATCTCCAAACCGGAATACTCAAAATGGAGCTATGATGACTTTTTTATAAG 904
D 166530 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166471
QY 905 GAATAAATACATACAGTCACTTATACAGAGGATTTAAAGATGGAGTGAACCAAGTTT 964
D 166470 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166411
QY 965 CAGCTGATTTACCAGRAAATACAAACATTCGACCATCAACCCCAATAGCAATGATG 1024
D 166410 ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 166351
QY 1025 ACCCAGGTCAAAAGTAAACAGTGGGATGGTCAAAATTAATTAATAAAAAAATGATGGT 1083
D 166350 ATACAGATACAGATACAGGATTTGGAGCTGGAGTCATTAAGAGGATCAAGAGTGCAGGT 166292

RESULT 16
AC137124 224391 bp DNA linear HTG 16-DEC-2003
LOCUS Mus musculus chromosome 3 clone RP24-229P11 map 3, *** SEQUENCING
DEFINITION IN PROGRESS ***, 7 unordered pieces.
ACCESSION AC137124
VERSION AC137124.4 GI:39930797
KEYWORDS HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 224391)
Birken, B., Nusbaum, C. and Lander, E.
MUS musculus chromosome 3, clone RP24-229P11
Unpublished
2 (bases 1 to 224391)
Birken, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
```

```
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Miho, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Noman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Petersen, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 224391)
Birken, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cooke, P., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Miho, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Petersen, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schnupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26589
Center clone name: 229_P_11
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 65156: contig of 65156 bp in length
* 65157 65256: gap of 100 bp
* 65257 67286: contig of 2030 bp in length
* 67287 67386: gap of 100 bp
* 67387 163968: contig of 96582 bp in length
* 163969 164068: gap of 100 bp
* 164069 189884: contig of 25816 bp in length
* 189885 189985: gap of 100 bp
* 189986 192533: contig of 2549 bp in length
```



[illegible]

polymerase II; FN transporter family; kinesin-related protein; p-type ATPase ring finger protein; PDZ domain; proteasome component C9; putative cleavage and polyadenylation specificity factor protein; putative inorganic pyrophosphatase; triosephosphate isomerase; trophozoite stage antigen; ubiquitin-conjugating enzyme; zinc-finger protein.

**SOURCE**  
Plasmodium falciparum 3D7

**ORGANISM**  
Plasmodium falciparum 3D7

**REFERENCE**  
1 (bases 1 to 164399)

**AUTHORS**  
Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T., Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T., Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S., Hornsby,T., Horrocks,P., Jagsels,K., Jassal,B., Kyes,S., McLean,J., Moulé,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A., Rajadream,M.-A., Rutter,S., Skelton,J., Squares,R., Squares,S., Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and Barrall,B.G.

**TITLE**  
The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum

**JOURNAL**  
Nature 400 (6744), 532-538 (1999)

**MEDLINE**  
99376085

**PUBMED**  
10448855

**REFERENCE**  
2 (bases 1 to 164399)

**AUTHORS**  
Lawson,D., Bowman,S. and Barrell,B.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (14-AUG-1997) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

**COMMENT**  
On or before May 14, 2001 this sequence version replaced gi:2665320, gi:2982570, gi:2982576, gi:2894503, gi:2982580, gi:2894373, gi:2982530.  
For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).

**FEATURES**  
Location/Qualifiers  
1..164399  
/organism="Plasmodium falciparum 3D7"  
/mol\_type="Genomic DNA"  
/isolate="3D7"  
/db\_xref="taxon:36329"  
/chromosome="3"  
/clone="MAL3P6"  
complement(152..1075)  
/gene="MAL3P6.1"  
/note="synonym: PFC0700c"  
complement(152..1075)  
/gene="MAL3P6.1"  
/note="CDS conserved in P. knowlesi and P. yoelii"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CABL1150.1"  
/db\_xref="GI:3758865"  
/db\_xref="SPTREMBL:O77394"  
/translation="MHDFFLKSKFNLSSPLFNFFYKRNEDEYFKKORNNDDDLGWMHYNADVEHREKHKDMRTSLKNEQELIYTYNNFNNNYNNKSTFKLXKNNED EYNEOEYETANLLQDKNKIINDMNNFYNNFNKSLNFONFQADKFKLYLLKNNK NNSTNENIADVEFKLKHVLFLQMNVLQKOLLTNHLINPKLMPHHIILNNKTEV SNQNVSEIQNNKDKKNGTMYLLKILSSRFNQMIFVSSIIISIFLYLNKHQWRAUKI SQQAKINSFNKSGVRLFEESLGRKNNKI"  
2866..3762  
/gene="PFC0701w"  
2866..3762  
/gene="PFC0701w"  
/note="Revised: new gene prediction; CDS conserved in P. knowlesi and P. yoelii"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB90508.1"  
/db\_xref="GI:7768284"  
/db\_xref="SPTREMBL:Q9NDU6"  
/translation="MNTKLFFGNSLCNKILYIKVNNLNRGCYNIAAGPFIISYLKLEYI NNRKBECCSSGNIIGKYGKNGKSPKFLKKNINFIYGNFNTCMIKNNYNNINII NSDIGNYVFTPKRYKMYKTSYKVKRSPNGOGAKIGIKRISGEVVRTGOMIYAV









**SOURCE** Danio rerio (zebrafish)  
**ORGANISM** Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
**REFERENCE** 1 (bases 1 to 200634)  
**AUTHORS** Mashreghi-Mohammadi, M.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (03-JUL-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
**COMMENT** On Jul 1, 2003 this sequence version replaced gi:3216885.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: zfish-help@sanger.ac.uk  
 -----  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em.: EMBL; Sw.: SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)  
 Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.  
 Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml) DKEY-12H12 is from a Zebrafish BAC library  
**VECTOR:** pIndicoBAC-5  
**FEATURES** Location/Qualifiers  
 source  
 1..200634  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-12H12"  
 /clone\_lib="DanioKey"  
**ORIGIN**  
 Query Match 5.1%; Score 77.4; DB 5; Length 200634;  
 Best Local Similarity 45.4%; Pred. No. 0.00067;  
 Matches 401; Conservative 0; Mismatches 471; Indels 11; Gaps 3;  
 QY 215 TAATTCCTCAAGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTATTTCTCTTTTA 274  
 DB 78947 TAAAGATCACAAATAAATGCTATTACTACTATCAATCAACTACTACTACTGATA 79006  
 QY 275 CGACAACACTAATGGAGGAGACATATGTAACTAAAAAGATAGCTGGTCAGCAATG 334  
 DB 79007 CTACTACTACTAATCAAACTACTACTACTATTAATAAATAAATACTACTACTA 79066  
 QY 335 AGATTCCGACATGGGCTAAATCTATATCACTAGCTAAATACACAGTTTCCACTGTTACTG 394  
 DB 79067 CTACTGCTAAAAA-ATATTACTATTACTCTCTTAATCAAACTACTACTGCTACTACTA 79125

QY 395 AGTCAATAATGATGGTACTGCTAGGTTTATTAATGTTTCCCAATATGGATATTATTATGTTT 454  
 DB 79126 CTATTACTAATCAAAATACTACTACTATTATTACTACTACTGCTAAAAAACTACTACTG 79185  
 QY 455 CTAGCACATGTTAATAATGGAGCTGTAATATGTTTACATCTGTAACTCCAAATGCTACTA 514  
 DB 79186 CTATTACTACTACTAATCAAACTACTACTACTAATTAATCAAAATACTGCTACTACTA 79245  
 QY 515 TTCATGAAAGAATACTGATCGGACATGGGGAGATGGTGGTGGAAAAAATCTGTAGATCAA 574  
 DB 79246 CTACTACTAATCAAACTACTACTACTAATCAAACTACTACTACTACTACTAATA 79305  
 QY 575 AAACGTACTCGTGGTGATACAGTCAAAATATATATATATTAAGAAATCGAGTCAATT 634  
 DB 79306 CTAATCAAAATACTGCTACTACTACTAATCAAACTACTACTACTAATCAAACTACTA 79365  
 QY 635 ATCATGGTACAGAAAAAGTGTATCAATATGTTTATAAAGGATACTATGTCATCTGCTCTG 694  
 DB 79366 AAACCTACTACTAATCAAACTACTACTACTAATCAAACTACTACTACTACTACTA 79425  
 QY 695 TAGTTGATTTGAACGAGAGGGTCTTATGAAGTAATCTATTACTGATGGATCAGGGAATATTA 754  
 DB 79426 CTATTAGTACTACTACTAGCACTATTA---TTCAAACTACTAATCAAACTACTACTACTA 79482  
 QY 755 CAACCTCTAACTCAAGTTCGGAAAAAGCACTGGGAAGTATAACCTGTTAGAGGAAAAATA 814  
 DB 79483 CTACTAATATTAATAATAATCAAACTACTACTACCACCTATTACTACTAACCAAACTACTA 79542  
 QY 815 ATAATTTCCAGCATACTATTTCGGTGGGAGCTACCAATCTCCCAACCGGAATACTCAAA 874  
 DB 79543 ACATTACTACTACTACTACTCAAACTACTACTACTAATCAAACTACTACTATTATCAAACTACTA 79602  
 QY 875 ATGGAGCTTAATGATGACATTTTTTTTATGAAGGAATAAATACAACTACAGCTACTATACAG 934  
 DB 79603 CTATTACTACTAATCAAACTACTACTACTACTAATTAATTAATCAATCAAACTACTACT 79662  
 QY 935 GAGTATTAAGAGTGGAGCTAAACAGGTTCCAGCTGATTTACCCAGAAAAATCAAAACATTTG 994  
 DB 79663 TACCACATTACTACTACTAACCAAACTA-----CTAACACTACTACTACTACTACTACTA 79715  
 QY 995 CGACCATCAACCCCAATACTAGCAATGATGACCCAGGTCAAAAGTAAACAGTACAGTGGGATG 1054  
 DB 79716 CTACTACTACTACTACTACTACTACTACTAATCAAACTACTACTAATCAAACTACTACT 79775  
 QY 1055 GTCAAAATTAATTAATAAATAATGATGTTCCACAAAAGCTTC 1097  
 DB 79776 CTACTATTACTACCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 79818  
**RESULT 21**  
**AL935062/c**  
**LOCUS** AL935062 176010 bp DNA linear VRT 30-AUG-2003  
**DEFINITION** Zebrafish DNA sequence from clone CH211-231P23 in linkage group 9, complete sequence.  
**ACCESSION** AL935062  
**VERSION** AL935062.13 GI:34368574  
**KEYWORDS** HTG.  
**SOURCE** Danio rerio (zebrafish)  
**ORGANISM** Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
**REFERENCE** 1 (bases 1 to 176010)  
**AUTHORS** Tromans, A.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (30-AUG-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
**COMMENT** On Aug 30, 2003 this sequence version replaced gi:32127703.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the cloning being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at

on the WormRep database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormrep](http://www.sanger.ac.uk/Projects/C_elegans/wormrep). Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr', 'drrr' and 'drrrr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see

Further information see  
[http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml)

[http://www.sanger.ac.uk/Projects/D\\_Ier18/lis](http://www.sanger.ac.uk/Projects/D_Ier18/lis)

CH211-231F23 1B 11C  
VECTOR: pTARBAC2.1

Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

## FEATURES

```

1. .176010
   /organism="Danio rerio"
   /mol_type="genomic DNA"
   /db_xref="taxon:7955"
   /clone="CH211-231F23"
   /clone_lib="CHORI-211"

```

## ORIGIN

Query Match 5.1%; Score 77.2; DB 5; Length 176010;  
Best Local Similarity 43.2%; Pred. No. 0.00075;  
Matches 367; Conservative 0; Mismatches 483; Indels 0; Gaps 0;

Qy	171	TAATGCAATGTATCTGATTCGCAATAAAGATGGAGCTTCTTATTTAAATTCCTCAAGGTAA	230
Db	80315	TATTACTACTACTACTAATAATAATAATAATTAATTATTATTATTATTATAATAC	80256
Qy	231	AGAAAGCTGAGTAAAGCTTCAACTGATTTTAAATTCCTTTTACGACCACTACTAATGG	290
Db	80255	CACGTGCTATTCTAAACACTACTACTACTACCAATACTACTCCACTACTACTACTACTAA	80196
Qy	291	AGGGAGAAATATGTAACTAAAAAAGATACCTGGTCAGCAAAATGAGATTGGACATGGGC	350
Db	80195	TACTACTACTACTATTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC	80136
Qy	351	TAAATCTATATCAGCTAAATACTACACAGGTTCCACTGTTACTGAGTCAAAATGATGG	410
Db	80135	TACTACTACTACTATTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAT	80076
Qy	411	TACTGAGGTTATTAAATGTTTCCCAATATGGATATTTATGTTTCTAGCACTGTTTAATAA	470
Db	80075	TACT	80016
Qy	471	TGGAGCTGTAATTTATGGTTTACATCTGTAATCTCCAAATGCTACTATTTCATGAAAAGAAATAC	530
Db	80015	TACTACTACTATTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	79956
Qy	531	TGATGGACATGGGGAGATGGTGGTGGAAAAAAGTGTAGATCAAAAACGTACTCGGTTGG	590
Db	79955	TAAATCTACTATTAAATAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	79896



/note="MLL-like"  
/rpt\_type=OTHER  
4455..4733  
/gene="Furu-I103"  
/number=3  
4734..4902  
/gene="Furu-I103"  
/number=3  
4903..5229  
/gene="Furu-I103"  
/number=4  
5230..5464  
/gene="Furu-I103"  
/number=4  
5465..5569  
/gene="Furu-I103"  
/number=5  
5570..5714  
/gene="Furu-I103"  
/number=5  
5715..5729  
/gene="Furu-I103"  
/number=6  
5730..5843  
/gene="Furu-I103"  
/number=6  
5844..5893  
/gene="Furu-I103"  
/number=7  
complement(7521..9972)  
/gene="psmb8"  
complement(join(7521..7609,7917..8112,8861..8990,  
9167..9278,9460..9601,9823..9972))  
/gene="psmb8"  
complement(join(7521..7609,7917..8112,8861..8990,  
9167..9278,9460..9601,9823..9972))  
/gene="psmb8"  
/codon\_start=1  
/product="low molecular mass polypeptide subunit PSMB8"  
/protein\_id="CAC13117.1"  
/db\_xref="GI:10803370"  
/db\_xref="GOA:Q9DES2"  
/db\_xref="SPTREMBL:Q9DES2"  
/translation="MALFDVSDTRYSELRLQILPSGQRLKERTTHYSGLKQFFA  
VPLGVDPMSYIKSCNNEAGVDIELDGTTLAPKFKHGVIVADSRASAGKYISSODA  
KXVEINSYLLGTWSSAADCNMYERLLAKECLRYLRNNHRLISVAAASKLLANMLG  
YRGMGLSVGSMICGWDKQLYYVDENGNRFSGQMFSGNSYAYAVDSGLREDWIVE  
EAYDLGRRGIVYATHRDAYSGGVYVNMHYHMMEDGWIKVCQDDVSQLIHLKYKEMF"  
complement(7521..7609)  
/gene="psmb8"  
/number=6  
complement(7610..7916)  
/gene="psmb8"  
/number=5  
complement(7917..8112)  
/gene="psmb8"  
/number=5  
complement(8113..8860)  
/gene="psmb8"  
/number=4  
complement(8861..8990)  
/gene="psmb8"  
/number=4  
complement(8991..9166)  
/gene="psmb8"  
/number=3  
complement(9167..9278)  
/gene="psmb8"  
/number=3  
complement(9279..9459)  
/gene="psmb8"  
/number=2  
complement(9460..9601)

intron  
/gene="psmb8"  
/number=2  
complement(9602..9822)  
/gene="psmb8"  
/number=1  
complement(9823..9972)  
/gene="psmb8"  
/number=1  
complement(10598..12501)  
/gene="psmb10"  
complement(join(10598..10709,10825..10976,11108..11166,  
11281..11396,11538..11678,11788..11885,12136..12223,  
12440..12501))  
/gene="psmb10"  
complement(join(10598..10709,10825..10976,11108..11166,  
11281..11396,11538..11678,11788..11885,12136..12223,  
12440..12501))  
/gene="psmb10"  
/codon\_start=1  
/product="low molecular mass polypeptide subunit PSMB10"  
/protein\_id="CAC13118.1"  
/db\_xref="GI:10803371"  
/db\_xref="GOA:Q9DES1"  
/db\_xref="SPTREMBL:Q9DES1"  
/translation="MALSNVLETAAGFNFDNAARNALRGLFEGGKTPKPMKTGTTI  
AGVVKDGVVLGADTRATSVADVADKCAKIHYPNIYCCGAGTAADTKTDLSS  
NLTFISLNSGRKPRVMAVNIQDTLYRHQIGIAPNLIIGVDCTGNHLKYKVPYGSV  
DKVPYLTWGSGLAALGILLEDGFKHMEVERATELRLAITHAGIMSLDGSNNIDICV  
ITRDVDYIRPYEESSEYKSRKTRYKYPGTPVLTKEKVPKLKLEMLQETVQRMDTV"  
complement(10598..10709)  
/gene="psmb10"  
/number=8  
complement(10710..10824)  
/gene="psmb10"  
/number=7  
complement(10825..10976)  
/gene="psmb10"  
/number=7  
complement(10977..11107)  
/gene="psmb10"  
/number=6  
complement(11108..11166)  
/gene="psmb10"  
/number=6  
complement(11167..11280)  
/gene="psmb10"  
/number=5  
complement(11281..11396)  
/gene="psmb10"  
/number=5  
complement(11397..11537)  
/gene="psmb10"  
/number=4  
complement(11538..11678)  
/gene="psmb10"  
/number=4  
complement(11679..11787)  
/gene="psmb10"  
/number=3  
complement(11788..11885)

Query Match 5.0%; Score 75.4; DB 5; Length 22398;  
Best Local Similarity 45.0%; Pred. No. 0.0028;  
Matches 329; Conservative 0; Mismatches 396; Indels 6; Gaps 1;  
Qy 172 AATGCAATGTATCTGATTCGAATAAGATGAGCTTCTTATTTAATCTCTCAAGTAAA 231  
Db 15368 AACGTCAATTTACATGATTTGGCGGATTACAGTTTACAGTACTTCTCTCTGTTGT 15427  
Qy 232 GAAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTCTTTTACGACAACTACTATGA 291  
Db 15428 GCATTTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTT 15487









QY 137 ATAAAGCATATAAAGTTTTTGATGAGAAATAGATAAATGCAAAATCTATCTGATTCGAATA 196  
Db |||||  
147414 ATAATAAATAAATAATGATAATAATAATAATAATAATGATGATAATAATAATAATA 147473  
QY 197 AAGATGGAGCTCTCTTTTAACTCTCAAGGTAAGAGAGCTGAGTATAAAGCTTCAACTG 256  
Db |||||  
147474 TTGATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 147533  
QY 257 ATTTTAAATCTCTTTTTACGACAACTACTAATGAGGAGGAGACATATGTAACATAAAAAAG 316  
Db |||||  
147534 ATAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATG 147593  
QY 317 ATACTGCGTCAGCAAAATGAGATGCGCATGGCTTAAATCTTATATCAGCTAATACTACAC 376  
Db |||||  
147594 ATAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 147653  
QY 377 CAGTTTCCACTGTTACTGAGTCAGTCAATATGATGGTACTGAGGTATTAATGTTTCCCAAT 436  
Db |||||  
147654 ATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 147713  
QY 437 ATGGATATTATATGTTTCTAGCACTGTTAATAATGAGAGCTGTAATTTATGGTTACATCTG 496  
Db |||||  
147714 ATAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 147773  
QY 497 TAACTCCAAATGCTACTATTTCATGAAAGAAATCTGATGCGGACATGGGAGATGGTGTG 556  
Db |||||  
147774 CTAATAATGACAATAATAATAATGACAATAATAATAATAATAATAATAATAATAATA 147833  
QY 557 GAAAACTGTAGTCAAAAAAGTACTCGTGGTGGTATGATACAGTCAAAATATACTATTCTT 616  
Db |||||  
147834 ATAATAATGATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 147893  
QY 617 ATAAGAATGCAGCTCAATATCATGTCAGAGAAAAGTGTATCAATGTATTATAAAGGAT 675  
Db |||||  
147894 ATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 147952

RESULT 26  
BX569779/c  
LOCUS  
DEFINITION  
BX569779  
ACCESSION  
BX569779.3 GI:35209937  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
Danio rerio (zebrafish)  
ORGANISM  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
REFERENCE  
1 (bases 1 to 183357)  
McLaren, S.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
JOURNAL  
On Sep 24, 2003 this sequence version replaced gi:33412519.  
COMMENT  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zC11319  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 181445 bases at least Q40  
Consensus quality: 181760 bases at least Q30  
Consensus quality: 182154 bases at least Q20  
Insert size: 182857; sum-of-ctigs  
Insert size: 176014; 6.0% error; agarose-fp

Quality coverage: 10.59x in Q20 bases; sum-of-ctigs Quality coverage: 11.06x in Q20 bases; agarose-fp

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 10212: contig of 10212 bp in length  
\* 10213: gap of 100 bp  
\* 10313: 38621: contig of 28309 bp in length  
\* 38622: gap of 100 bp  
\* 38722: 87713: contig of 48992 bp in length  
\* 87714: gap of 100 bp  
\* 87814: 91365: contig of 3552 bp in length  
\* 91366: gap of 100 bp  
\* 91466: 157405: contig of 65940 bp in length  
\* 157406: gap of 100 bp  
\* 157506: 161885: contig of 4380 bp in length  
\* 161886: gap of 100 bp  
\* 161986: 170778: contig of 8793 bp in length  
\* 170779: gap of 100 bp  
\* 170879: 183357: contig of 12479 bp in length.

FEATURES

source  
1. 183357  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clones="CH211-11319"  
/clone\_lib="CHORI-211"  
1. 10212  
/note="assembly fragment:00930"  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:left  
10313..38621  
/note="assembly fragment:01719"  
fragment\_chain:1  
38722..87713  
/note="assembly fragment:02253"  
fragment\_chain:1  
87814..91365  
/note="assembly fragment:01340"  
fragment\_chain:2  
91466..157405  
/note="assembly fragment:02753"  
fragment\_chain:2  
157506..161885  
/note="assembly fragment:01368"  
fragment\_chain:3  
161986..170778  
/note="assembly fragment:00374"  
fragment\_chain:3  
170879..183357  
/note="assembly fragment:00406"  
clone\_end:T7  
vector\_side:right

ORIGIN

Query Match 5.0%; Score 75; DB 2; Length 183357;  
Best Local Similarity 45.8%; Pred. No. 0.0017;  
Matches 258; Conservative 0; Mismatches 305; Indels 0; Gaps 0;  
QY 351 TAAATCTATATCAGCTAATACTACACCAGTTTCCACTGTTACTGAGTCAATAATATGATGG 410  
Db |||||  
140378 TAATAGTAAAAAATGCTGATAATAATAATAATAATCATCATCATCATTTTAATAATAATA 140319  
QY 411 TACTGAGGTTTAAATGTTTCCCAATATGATATATATGTTTCTAGCACTGTTAATAA 470  
Db |||||  
140318 TAATAATAATAATAGTAAATAATAAGGATGATGATGATAATAATAATAATAATAACAA 140259

\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

Accession	Contig	Size (bp)
19051	contig of 19051 bp in length	
19131	gap of 100 bp	
19152	contig of 19152 bp in length	
19159	contig of 3539 bp in length	
21691	gap of 100 bp	
21791	contig of 4788 bp in length	
26578	gap of 100 bp	
26579	contig of 19064 bp in length	
45742	contig of 100 bp	
45743	gap of 100 bp	
45842	gap of 100 bp	
45843	contig of 4282 bp in length	
50124	contig of 100 bp	
50125	gap of 100 bp	
50225	contig of 28061 bp in length	
78285	contig of 100 bp	
78385	gap of 100 bp	
78386	contig of 5854 bp in length	
84239	contig of 100 bp	
84240	gap of 100 bp	
84339	contig of 95823 bp in length	
180162	contig of 100 bp	
180262	gap of 100 bp	
182961	contig of 2699 bp in length	
182962	contig of 100 bp	
183061	gap of 100 bp	
183062	contig of 15637 bp in length	

FEATYPES	Location/Qualifiers
183002	130030: CONCRETS

source

```
1. .199698
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-222F23"
/clone_lib="CHORI-211"
```

1. 19051

```
1. :12051
/note="assembly fragment:02835
```

```
/IOCC= assembly_1_1"
fragment chain:1"
```

19152. 21690

```
1512: 02440
/note="assembly fragment: 02440
```

```
/proc= assembly 1"
fragment chain: 1"
```

21791. .26578

```
21751: 40570
/note="assembly fragment: 00503
```

```

/molecule assembly 1
fragment chain: 1"

```

Fragment Chair. 1  
26679 45742

```
20073. : 45742
/note="assembly fragment:00215
```

```
fragment chain: 2"
/molecule=assembly 1
```

Diagnostic Chart: 2  
45843 50124

```
43843: .30124
/note="assembly fragment: 02548
```

```
/note="assembly_1"  
fragment chain: 2"
```

```

fragment_chain:Z"
5032F 78285

```

```
50225. : 78285  
/proto="paccmbly" fragment: 01134
```

```
/note="assembly_1"
```

```
fragment_chain:3"
70200 64220
```

78386. .84239  
/rate="assembly" element="02498

```
/note="assembly_1"
```

```
fragment_chain:3"
00000000
```

84340. .180162

```
/note="assembly_
```

180263. .182961

```
/note="assembly_fragment:02452"
```

183062. .199698

## ORIGIN

Query Match	4.9%;	Score 74;	DB 2;	Length 199698;
Best Local Similarity	44.3%;	Pred. No. 0.0024;		
Matches 302. Conservative		0: Mismatches 380;	Indels 0;	Gaps 0;

137 ATAAAGCATATAAAGTTTTTGATGCAGAAATAGATAATGCAAATGTATCTGATTCGAATA 196

[illegible]

716 ATATGAAGAAGATGATGATGAAGGATGAAGATGATGATGAAGAGAACGATG 657

[illegible]

-----  
 \* NOTE: This is a 'working draft' sequence. It currently



```
/translation="MEEQNEEIEKTVSESEINEEVEBELFELPASSNDLVKKSRLI
ILSLLEKEFIDHIIKNNKKYKTLIISLNSRCKLLTLECLINLSAQPKELTERNLI
EILFDMKEEIEEDNCIDLYIMIISNLTCKEGYKVLDDINNSNINIKEDNFKVSF
FLNKLLYFPFLPKPSINKNLSKDYIYVSHVLINISIKESIVFKNVAFLNKISDQI
LNVESFRALIPFIINCLINEVTHPYIHDCCVLPFVLSYLVYNDYNTIKSGSYNNSN
NNEENTQNIHIIWNKSSILVSCSVIKSRILILFYLCDRDYSRNDIENVTKNLIEG
SNSRAIAI"
repeat_region 4677..4714
/rpt_type=tandem
repeat_region 5600..5641
/rpt_type=tandem
repeat_region 5745..5791
/rpt_type=tandem
repeat_region 6108..6152
/rpt_type=tandem
repeat_region 6168..6222
/rpt_type=tandem
repeat_region 6220..6268
/rpt_type=tandem
repeat_region 6420..6481
/rpt_type=tandem
repeat_region 6563..6621
/rpt_type=tandem
repeat_region 6718..6780
/rpt_type=tandem
repeat_region 6719..6790
/rpt_type=tandem
repeat_region 7543..7568
/rpt_type=tandem
repeat_region 7761..7816
/rpt_type=tandem
repeat_region 7824..7849
/rpt_type=tandem
repeat_region 7871..7930
/rpt_type=tandem
gene complement(<7953..>8438)
mRNA locus_tag="PF14_0240"
CDS complement(<7953..>8438)
locus_tag="PF14_0240"
/codon_start=1
/product="ribosomal protein L21e, putative"
/db_xref="GI:23497308"
/translation="MGKSKGSKSGTYYKFSKFRKHGECTANKYLEKLNVDYDVI
CDSTQQGMFPNYTHGKIGKLFHITKRGVGLNKRVRKHIEQKKVCVLEHVRSRC
NEDFLLRKIKNAELIKEALKUNEHINIKRKTEGPKPAAAIKVPSPKIIITIEPLPFE
Y"
repeat_region 8467..8550
/rpt_type=tandem
repeat_region 8473..8539
/rpt_type=tandem
repeat_region 8477..8537
/rpt_type=tandem
repeat_region 8723..8844
/rpt_type=tandem
repeat_region 8723..8840
/rpt_type=tandem
repeat_region 8782..8844
/rpt_type=tandem
repeat_region 9019..9059
/rpt_type=tandem
repeat_region 9019..9058
/rpt_type=tandem
repeat_region 9133..9189
/rpt_type=tandem
repeat_region 9221..9254
/rpt_type=tandem
repeat_region 9604..9637
/rpt_type=tandem
repeat_region 9668..9723
/rpt_type=tandem
```

```
repeat_region 10036..10073
/rpt_type=tandem
repeat_region 10445..10472
/rpt_type=tandem
repeat_region 10511..10551
/rpt_type=tandem
gene complement(<10565..>11080)
mRNA locus_tag="PF14_0241"
CDS complement(<10565..>11080)
locus_tag="PF14_0241"
/codon_start=1
/product="basic transcription factor 3b, putative"
/db_xref="GI:23497309"
/translation="MEKISPEILAAARAKLEKMGNNLROIGGKSGARRKIKKVKHNSI
SNEKKINIILKKIGASYFGDVDEICVTRTGDTFLEPKPKLCLASLOSNTYIVTGKNE
HKIDINKLFEGLKGNKNLDNMNLEKINIKNDPNILNKESGDTPKREDEQEAANDVPD
LVNFERVSKS"
repeat_region 11429..11458
/rpt_type=tandem
repeat_region 11514..11584
/rpt_type=tandem
repeat_region 11514..11579
/rpt_type=tandem
Query Match 4.9%; Score 73.8; DB 3; Length 250029;
Best Local Similarity 46.3%; Pred. No. 0.0024;
Matches 243; Conservative 0; Mismatches 282; Indels 0; Gaps 0;
QY 156 TGATGCAGAAATAGATAATGCAAAATGATCTGATTCGAAATAAGATGGAGCTTCTTATTT 215
Db 48791 TCATAATAATAATGATAATAATAATAATGATAATGATAATAATAATAATAATAATAATA 48850
QY 216 AATTCCTCAAGSTAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTAC 275
Db 48851 TAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 48910
QY 276 GACAACTACTAATGGAGGGAGAACATATGTAACATAAAAGATACGCGTCAGCAATGA 335
Db 48911 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 48970
QY 336 GATTCGGACATGGCTAAATCTATATCAGCTAAATACTACACAGTTTCCACTGTTACTGA 395
Db 48971 TAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49030
QY 396 GTCAATAATGATGCTACTGAGGTTTATTAATGTTTCCCAATATGGATATTTATGTTTC 455
Db 49031 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49090
QY 456 TAGCCTGTTAATAATGGAGCTGTAATTTATGTTTACATCTGTACTCCAAATGCTACTAT 515
Db 49091 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49150
QY 516 TCATGAAGAAGTACTGATCGACATGGGAGATGGTGTGGAAAACTGTAGATCAAAA 575
Db 49151 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49210
QY 576 AAGCTACTCGTTGGTGATACAGTCAAAATATATCTATTACTTTAAGAGATGCACTCAATTA 635
Db 49211 CAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49270
QY 636 TCATGTACAGAAAAGTGTATCAATATGTTTATAAGGATCTAT 680
Db 49271 TAATAACAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATGT 49315
```

RESULT 29  
AF413050S2/c AF413050S2 666 bp DNA linear PLN 31-DEC-2001  
LOCUS Zea mays CAG and TAG long trinucleotide repeat.  
DEFINITION AF413051  
ACCESSION AF413051  
VERSION AF413051.1 GI:18001263

KEYWORDS	2 of 2	LOCUS	BX321884	199347 bp	DNA	linear	HTG 24-SEP-2003
SEGMENT	Zea mays	DEFINITION	Danio rerio clone DKEYP-29A10, WORKING DRAFT SEQUENCE, 12 unordered pieces.				
SOURCE	Zea mays	ACCESSION	BX321884				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.	VERSION	BX321884.5	GI:35209510			
		KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.				
REFERENCE	1 (bases 1 to 666)	SOURCE	Danio rerio (zebrafish)				
AUTHORS	Ananiev,E.V., Lorentzen,J. and Bruggemann,E.	ORGANISM	Danio rerio				
JOURNAL	Microsatellite megatracts in the maize (Zea mays L.) genome		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.				
REFERENCE	2 (bases 1 to 666)	REFERENCE	1 (bases 1 to 199347)				
AUTHORS	Ananiev,E.V., Lorentzen,J. and Bruggemann,E.	AUTHORS	McLaren,S.				
JOURNAL	Direct Submission	JOURNAL	Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk				
FEATURES	Avenue, Johnston, IA 50331, USA	COMMENT	On Sep 24, 2003 this sequence version replaced gi:3260915.				
source	Location/Qualifiers		----- Genome Center				
	1..666		Center: Wellcome Trust Sanger Institute				
	/organism="Zea mays"		Center code: SC				
	/mol_type="genomic DNA"		Web site: http://www.sanger.ac.uk				
	/db_xref="taxon:4577"		Contact: zfish-help@sanger.ac.uk				
misc_feature	1..666		----- Project Information				
repeat_region	/note="3' end of a 2.6 kbp fragment"		Center project name: zkp29A10				
	1..537		----- Summary Statistics				
	/note="CAG and TAG long trinucleotide repeat; organized in clusters"		Assembly program: XGAP4; version 4.5				
	/rpt_type=tandem		Chemistry: Dye-terminator; 100% of reads				
	/rpt_unit="cag;tag"		Consensus quality: 195192 bases at least Q40				
misc_feature	538..566		Consensus quality: 196456 bases at least Q30				
	/note="similar to LTR of Zeon 1 retrotransposon of Zea mays"		Consensus quality: 197320 bases at least Q20				
			Insert size: 198247; sum-of-contigs				
ORIGIN			Insert size: 172014; 5.6% error; agarose-fp				
Query Match	4.8%; Score 73; DB 8; Length 666;		Quality coverage: 7.64x in Q20 bases; sum-of-contigs Quality coverage: 10.37x in Q20 bases; agarose-fp				
Best Local Similarity	47.3%; Pred. No. 0.021;						
Matches 220; Conservative 0; Mismatches 245; Indels 0; Gaps 0;							
QY 100 ACAATTACAGTTCAAGATACATAAAAGCGCAACTATAAAGCATATAAAGTTTTGAT 159							
DB 540 ACAACTACTACTGTTACTACTACTACTACTACTACTACTACTACTACTACTACT 481							
QY 160 GCAGAAATAGATAGCAATGTATCTGATTCGATTAAGATGAGGAGCTTCTTTAATT 219							
DB 480 ACTGCTACTGCTACTACTACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGTT 421							
QY 220 CCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTACGACA 279							
DB 420 GCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 361							
QY 280 ACTACTAATGGAGGAGAACATATGTAATAAAGATACCTGCTGCTGCTGCTGCTGCTGCT 339							
DB 360 ACTACTGCTGCTACTACTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 301							
QY 340 GCGACATGGCTAAATCTATATCAGTAATATACACAGTTTCCACTGTTACTGAGTCA 399							
DB 300 GCTACTCTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 241							
QY 400 AATAATGATGTTACTGAGGTAAATGTTTCCCAATATGATATATTATGTTTCTAGC 459							
DB 240 GCTACTGCTGCTACTACTGCTACTGCTACTGCTACTGCTACTACTACTACTACTACTACT 181							
QY 460 ACTGTTAATATGGAGCTGTAATATGTTAGTTACATCTGTAACCTCCAAATGCTACTATTGAT 519							
DB 180 GCTACTGCTACTGCTACTACTACTATTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 121							
QY 520 GAAAGAAATACATGATCGGACATGGGAGATGGTGGTGGGAAACT 564							
DB 120 GCTACTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 76							
FEATURES							
source							
	1..199347						
	/organism="Danio rerio"						
	/mol_type="genomic DNA"						
	/db_xref="taxon:7955"						
	/clone="DKEYP-29A10"						
RESULT 30							
BX321884							

misc_feature	/clone_lib="DanieKeypilot"	
	1..94733	
misc_feature	/notes="assembly_fragment:00836"	
	fragment_chain:1	
misc_feature	94834..139839	
	/notes="assembly_fragment:00061"	
misc_feature	fragment_chain:1	
	139940..142547	
misc_feature	/notes="assembly_fragment:00422"	
	142648..145652	
misc_feature	/notes="assembly_fragment:00804"	
	145753..147869	
misc_feature	/notes="assembly_fragment:00868"	
	147970..150959	
misc_feature	/notes="assembly_fragment:01524"	
	151060..154357	
misc_feature	/notes="assembly_fragment:01995"	
	154458..156995	
misc_feature	/notes="assembly_fragment:02125"	
	157096..159488	
misc_feature	/notes="assembly_fragment:02132"	
	159589..161652	
misc_feature	/notes="assembly_fragment:02163"	
	161753..165174	
misc_feature	/notes="assembly_fragment:02166"	
	163275..199347	
misc_feature	/notes="assembly_fragment:02257.0"	

ORIGIN

Query Match 4.8%; Score 73; DB 2; Length 199347;

Best Local Similarity 43.9%; Pred. No. 0.0035;

Matches 446; Conservative 0; Mismatches 565; Indels 4; Gaps 3;

Qy	110	TTCAAGTACTCAAAAGGCGCACTATAAGCATATAAGCTTTTGATGCGAATAAG 169
Db	168472	TTGTATGTAGTTTCAAAATGAAATTTATTACTCAGACATTTATTGCTTTTAAATAATAATA 168531
Qy	170	ATATGCAATGTATCTGATTGGAATAAGATGGAGCTTCTTATTTAAATTCCTCAAGGTA 229
Db	168532	ATA 168591
Qy	230	AAGAAGCTGAGTATAAGCTTCAACTGATTTAATTCCTCTTTTACGACAACTACTAATG 289
Db	168592	ATA 168651
Qy	290	GAGGAGAACATATCTACTAAAGATCTGCGTCAGCAATGAGATTGCGACATGGG 349
Db	168652	ATAATAGTAGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 168711
Qy	350	CTAAATCTATACAGTAAATACTACACCAGTTTCCACTGTTTACTGAGTCAAAATAATGATG 409
Db	168712	GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 168771
Qy	410	GTACTGAGGTTATTAATGTTTCCCAATAGGATATTTATGTTTCTAGCACTGTTAATA 469
Db	168772	ATAATAATAGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 168831
Qy	470	ATGGAGCTGTAATTTGTTTACATCTGTAATCTCAAAATGCTACTATTCTATGAAAGAAATA 529
Db	168832	GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 168891
Qy	530	CTGATGCGACATGGGAGATGGTGGGAAACCTGTAGATCAAAACAGTACTCGGTTG 589
Db	168892	GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 168951
Qy	590	GTGATACAGTCAAAATATACTATTACTTATAGATGCGAGTCAATTTATCATGGTACAGAAA 649
Db	168952	TAAATAAAAGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 169011
Qy	650	AAAGTGATCAATATGTTATAAGGATCTATGCCATCTGCTTCTGTAGTTGATTTGAACG 709
Db	169012	GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 169071

Search completed: July 27, 2004, 00:04:31

Job time : 6040 secs

GenCore version 5.1.6  
 >Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 00:07:57 ; Search time 666 Seconds  
 (without alignments)  
 9625.422 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaaagaataatttcac.....gtcgtcgcttactgtcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : N\_Geneseq\_29Jan04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1509	100.0	1509	5	AAD02390 Virulent
2	1509	100.0	1509	6	ABK11585 S. agalac
3	64.8	4.3	1983	2	AAX20108 Enterococ
4	64.8	4.3	1983	6	ABN98093 E faecali
5	64.8	4.3	1983	7	ACA88057 E. faecal
6	64.8	4.3	1983	7	ABX61663 Enterococ
7	64.8	4.3	2199	6	ABK11591 E. faecal
8	64.8	4.3	15614	2	AAX12982 Enterococ
9	64.8	4.3	15614	6	ABN98777 Enterococ
10	59.6	3.9	11922	7	AAX70187 Plasmodi
11	59.2	3.9	1137	7	ACA28118 Prokaryot
12	56.6	3.8	969	6	ABQ39490 Oligonuc
13	56.6	3.8	969	6	ABQ39491 Oligonuc
14	55.4	3.7	2408	4	ABL28164 Drosophil
15	55.2	3.7	32392	6	ABL56203 AmEPV gen
16	55.2	3.6	408	4	ABL28165 Drosophil
17	53.6	3.6	2535	4	ABL25476 Drosophil
18	53	3.5	4985	6	ABQ75107 Anopheles
19	53	3.5	4985	9	ACF79720 Mosquito
20	53	3.5	6033	3	AAA70152 Plasmodi
21	51.2	3.4	2703	6	ABN67916 Streptoco
22	51.2	3.4	110000	6	ABN71527_13
23	50.4	3.3	14066	2	AAX99556 Nucleic a

RESULT 1

24	50.4	3.3	14067	7	ACA52811	ACA52811 Prokaryot
25	50.2	3.3	7143	3	AAA70250	Plasmodi
26	50	3.3	3996	6	AAD47004	Plasmodi
27	49.8	3.3	1687	2	AAX20109	Enterococ
28	49.8	3.3	1687	6	ABN98094	E faecali
29	49.8	3.3	1687	7	ACA88058	E. faecal
30	49.8	3.3	1687	7	ABX61664	Enterococ
31	49.6	3.3	3837	3	AAA70211	Plasmodi
32	48.6	3.2	4997	7	ACA47636	Prokaryot
33	48.6	3.2	4851	6	AAS20800	Clostridi
34	48.4	3.2	4197	2	RAQ99430	B. sphaer
35	47.8	3.2	740	4	AAC85918	ICP41 cDN
36	47.8	3.2	1083	5	AAS76745	DNA encod
37	47.8	3.2	8244	5	AAS66529	DNA encod
38	47.6	3.2	11143	4	ABL12834	Drosophil
39	47.4	3.1	1037	3	AAA59242	Exons E.
40	47.4	3.1	1472	3	AAA59241	Exons D.
41	47.4	3.1	2428	7	ADA89818	Staphyloc
42	47.4	3.1	3945	2	AAX77593	S. aureus
43	47.4	3.1	4047	4	AAS52261	Staphyloc
44	47.4	3.1	4047	7	ACF73996	Staphyloc
45	47.4	3.1	4050	4	AAS55402	Staphyloc
46	47.4	3.1	4143	7	ABT14918	Pathogen
47	47.4	3.1	4418	2	AAV74765	Staphyloc
48	47.4	3.1	7458	3	AAA70106	Plasmodi
49	47.2	3.1	5314	9	ADB46161	Plasmodi
50	47	3.1	3567	3	AAA70117	Plasmodi
51	47	3.1	5940	3	AAA70105	Plasmodi
52	46.8	3.1	1938	4	AAS53795	S. epider
53	46.8	3.1	1948	2	AAQ12693	Toxin A g
54	46.8	3.1	2801	7	ACA46974	Prokaryot
55	46.8	3.1	3188	4	AAH54473	S. epider
56	46.8	3.1	3594	4	AAH54439	S. epider
57	46.8	3.1	3641	4	AAH54487	S. epider
58	46.8	3.1	4506	4	AAH53728	S. epider
59	46.8	3.1	4590	6	ABN91645	Staphyloc
60	46.8	3.1	7047	6	ABK28386	DNA trans
61	46.2	3.1	535	6	ABQ42292	Oligonuc
62	46.2	3.1	535	6	ABQ42293	Oligonuc
63	46.2	3.1	4158	7	ACA20270	Prokaryot
64	46	3.0	2394	9	ADC89717	L. johnso
65	46	3.0	2900	9	ADC89716	L. johnso
66	46	3.0	5139	3	AAA70139	Plasmodi
67	45.8	3.0	1159	3	AAA59240	An EcoRI
68	45.8	3.0	110000	6	ABQ67196_1	Continuation (2 of
69	45.8	3.0	110000	6	ABQ69245_26	Continuation (27 o
70	45.6	3.0	2017	9	ADB46163	Plasmodi
71	45.6	3.0	4423	2	AAQ29470	Extracell
72	45.6	3.0	6397	4	AAS45400	Chemical
73	45.6	3.0	6397	4	AAS46459	Tumour su
74	45.6	3.0	6397	6	ABL33346	Human imm
75	45.6	3.0	6397	6	ABK28245	DNA trans
76	45.6	3.0	6744	2	AAQ29471	Extracell
77	45.6	3.0	50000	6	ABL56202	AmEPV gen
78	45.4	3.0	2217	7	ACA29569	Prokaryot
79	45.4	3.0	2505	6	ABN66950	Streptoco
80	45.4	3.0	3927	3	AAA70101	Plasmodi
81	45.4	3.0	110000	6	ABN71527_19	Continuation (20 o
82	45.2	3.0	861	6	ABL55632	AmEPV tri
83	45.2	3.0	3931	8	ACB84661	P. falcip
84	45.2	3.0	50000	6	ABL56201	AmEPV gen
85	45.2	3.0	110000	6	ABA90521_14	Continuation (15 o
86	45	3.0	12237	6	ABL34358	Human imm
87	44.8	3.0	16287	6	ABL32672	Human imm
88	44.6	3.0	954	6	ABQ76556	C. albica
89	44.6	3.0	11679	8	ADA30090	DNA encod
90	44.4	2.9	1998	3	AAA70212	Plasmodi

ALIGNMENTS



```

AAD02390
ID AAD02390 standard; DNA; 1509 BP.
XX
AC AAD02390;
XX
DT 24-APR-2001 (first entry)
XX
DE Virulent group B Streptococcus agalactiae spb1 DNA.
XX
KW Type III virulent group B; spb1; cell wall bound protein; antibacterial;
KW immunisation; group B streptococci; GBS infection; vaccine; bacteraemia;
KW pneumonia; meningitis; endocarditis; osteoarticular infection; ds.
XX
OS Streptococcus agalactiae.
XX
PH Key Location/Qualifiers
FT CDS 1..1509
FT /*tag= a
FT /product= "S. agalactiae spb1 protein"
FT sig_peptide 19..87
FT /*tag= b
FT mat_peptide 88..1506
FT /*tag= c
FT /product= "S. agalactiae mature spb1 protein"
XX
XX WO20007878-A1.
XX
XX 28-DEC-2000.
XX
XX 21-JUN-2000; 2000WO-US017082.
XX
XX 21-JUN-1999; 99US-0140084P.
XX
XX (UTAH ) UNIV UTAH RES FOUND.
XX
XX Adderson E, Bohnsack J;
XX
XX WPI; 2001-102693/11.
XX
XX P-PSDB; AAY72357.
XX
XX Polynucleotide from spb1 and 2 genes derived from virulent Group B
XX streptococci, polypeptide encoded by the polynucleotide useful as vaccine
XX for immunizing a mammal against the streptococcal infection.
XX
XX Claim 9; Page 20-23; 34pp; English.
XX
XX The present sequence is spb1 gene from type III virulent group B
XX Streptococcus agalactiae. The spb1 protein has the characteristics of a
XX cell wall bound protein and has antibacterial activity. The N-terminus
XX of the spb1 protein is a hydrophilic, basic stretch of 6 amino acids
XX followed by a 23 amino acid hydrophobic, proline rich core, consistent
XX with a signal peptide. The hydrophilic mature protein terminates in
XX atypical LPXTG domain that immediately precedes a hydro- phobic 20 amino
XX acid core and a short, basic hydrophilic terminus. The spb1 protein is
XX used as a vaccine to immunise mammals against group B streptococci (GBS)
XX infection (e.g. bacteraemia, pneumonia, meningitis, endocarditis and
XX osteoarticular infections). Determination of the gene products specific
XX to type III-3 GBS is useful for diagnosing mammals infected or colonised
XX by virulent GBS
XX
XX Sequence 1509 BP; 534 A; 241 C; 298 G; 436 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 1509; DB 5; Length 1509;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGAAAAGAAATGATTCAATCGCTGTAGTCGCGAGTTTACATTTGGTATGGCTGTA 60
XX |||||||
XX 1 ATGAAAAGAAATGATTCAATCGCTGTAGTCGCGAGTTTACATTTGGTATGGCTGTA 60
XX |||||||
XX 61 TCACCAGTTACGCCGATAGCTTTTGGCCGCTGAGACAGGACAATTACAGTTCAAGATACT 120
XX |||||||
XX 61 TCACCAGTTACGCCGATAGCTTTTGGCCGCTGAGACAGGACAATTACAGTTCAAGATACT 120
XX |||||||

121 CAAAAGCGCAACCTATAAGCATATAAAGTTTTTGTGACAGAAATAGATAAGTCAAAAT 180
121 CAAAAGCGCAACCTATAAAGCATATAAAGTTTTTGTGACAGAAATAGATAAGTCAAAAT 180
181 GTATCTGATTGCAATAAAGATGGAGCTTCTTATTAAATTCCTCAAGGTAAGAAGCTGAG 240
181 GTATCTGATTGCAATAAAGATGGAGCTTCTTATTAAATTCCTCAAGGTAAGAAGCTGAG 240
241 TATAAGCTTCAACTGATTTTAAATTTCTTTTACGACAACTACTATAATGGAGGAGAA 300
241 TATAAGCTTCAACTGATTTTAAATTTCTTTTACGACAACTACTATAATGGAGGAGAA 300
301 TATGTAACTAAAAAGATATCTCGCTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATA 360
301 TATGTAACTAAAAAGATATCTCGCTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATA 360
361 TCAGCTAATATACACCAAGTTTCCACTGTTACTGAGTCAAAATATGATGGTACTGAGTT 420
361 TCAGCTAATATACACCAAGTTTCCACTGTTACTGAGTCAAAATATGATGGTACTGAGTT 420
421 ATTAATGTTTCCCAATATGGATATTTATTTCTAGCACTGTTTAATAATGGAGCTGTA 480
421 ATTAATGTTTCCCAATATGGATATTTATTTCTAGCACTGTTTAATAATGGAGCTGTA 480
481 ATTAATGTTTCCCAATATGGATATTTATTTCTAGCACTGTTTAATAATGGAGCTGTA 540
481 ATTAATGTTTCCCAATATGGATATTTATTTCTAGCACTGTTTAATAATGGAGCTGTA 540
541 TGGGAGATGGTGGGAAAGCTCTAGATCAAAAAGCTACTCGGTTGGTATACAGTC 600
541 TGGGAGATGGTGGGAAAGCTCTAGATCAAAAAGCTACTCGGTTGGTATACAGTC 600
601 AAATATATCTATTACTTATAAGAAATGCAGTCAATTTATCATGTTACGAAAGAAAGTGTATCAA 660
601 AAATATATCTATTACTTATAAGAAATGCAGTCAATTTATCATGTTACGAAAGAAAGTGTATCAA 660
661 TATGTTATAAGGATATATGCGCATCTGCTCTGTAGTTGATTTCGAAAGGAGTCTTAT 720
661 TATGTTATAAGGATATATGCGCATCTGCTCTGTAGTTGATTTCGAAAGGAGTCTTAT 720
721 GAAGTAACTATTACTGATGGATCAGGGAATATTACAACCTCAAGTCAAGGTTTCGAAAAA 780
721 GAAGTAACTATTACTGATGGATCAGGGAATATTACAACCTCAAGTCAAGGTTTCGAAAAA 780
781 GCAACTGGGAAGTATAAAGCTTTAGAGGAAATAATAATTTCAAGTACTATTCGCTGG 840
781 GCAACTGGGAAGTATAAAGCTTTAGAGGAAATAATAATTTCAAGTACTATTCGCTGG 840
841 GCAGCTACCAATATCTCCACCGGAATATCTCAAAATGGAGCTAAATGATGACTTTTTTAT 900
841 GCAGCTACCAATATCTCCACCGGAATATCTCAAAATGGAGCTAAATGATGACTTTTTTAT 900
901 AAGGGAATAAATAACATCAAGTCACTTATACAGGATATTAAAGAGTGGAGCTAAACCA 960
901 AAGGGAATAAATAACATCAAGTCACTTATACAGGATATTAAAGAGTGGAGCTAAACCA 960
961 GGTTCAGCTGATTTACAGAAATCAACAATGCGAATCAATCAATCAATCAATCAATCAAT 1020
961 GGTTCAGCTGATTTACAGAAATCAACAATGCGAATCAATCAATCAATCAATCAATCAAT 1020
1021 GATGACCCAGGTCAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1080
1021 GATGACCCAGGTCAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1080
1081 GGTTCACAAAAAGCTTCATTACAAGTGCTATATTTGTTTTTAAAGAAATGCTACGGGTCAA 1140
1081 GGTTCACAAAAAGCTTCATTACAAGTGCTATATTTGTTTTTAAAGAAATGCTACGGGTCAA 1140
1141 TTTCTAAACTTTAACGATACAAATTAACGTTGAATGGGCGACAGAGCTTAATGCAACAGAA 1200
1141 TTTCTAAACTTTAACGATACAAATTAACGTTGAATGGGCGACAGAGCTTAATGCAACAGAA 1200
```



QY 1201 TATACACAGGAGCAGATGGTATTAATTACCATTCACAGCGTTGAAAGAGGTACATCTAT 1260  
DB 1201 TATACACAGGAGCAGATGGTATTAATTACCATTCACAGCGTTGAAAGAGGTACATCTAT 1260  
QY 1261 CTAGTTGAGAAAAGGCTCCCTTAGCTTACAAATTTGTTAGATAACTCTCAGAAGGTTATT 1320  
DB 1261 CTAGTTGAGAAAAGGCTCCCTTAGCTTACAAATTTGTTAGATAACTCTCAGAAGGTTATT 1320  
QY 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTTAGTTAAACCCAACTGTT 1380  
DB 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTTAGTTAAACCCAACTGTT 1380  
QY 1381 GAAATTAACAAGGCTACTGAGTTGGCTTCAACAGGTGGTATTTGGTACAAATTTTCTAC 1440  
DB 1381 GAAATTAACAAGGCTACTGAGTTGGCTTCAACAGGTGGTATTTGGTACAAATTTTCTAC 1440  
QY 1441 ATTATAGGTGCAATTTTACTAATAGCAGCAGGTATCTGCTGTTGCTCGTCTGTTTA 1500  
DB 1441 ATTATAGGTGCAATTTTACTAATAGCAGCAGGTATCTGCTGTTGCTCGTCTGTTTA 1500  
QY 1501 CGTTCTTAA 1509  
DB 1501 CGTTCTTAA 1509  
RESULT 2  
ABK11585  
ID ABK11585 standard; DNA; 1509 BP.  
AC ABK11585;  
XX  
DT 05-JUN-2002 (first entry)  
DE S. agalactiae Sbp1 gene.  
XX  
KW Extracellular matrix adhesion; Ema; ds; group B streptococcus; GBS;  
KW DNA vaccine; Sbp1; Sbp2; Rib; Lmb; Csa-ase; C protein alpha antigen;  
KW neonatal bacterial infection; gene.  
XX  
OS Streptococcus agalactiae.  
XX  
FH Key Location/Qualifiers  
FT 1..1509  
FT /\*tag= a  
FT /\*product= "Sbp1"  
XX  
PN WO200212294-A2.  
XX  
PD 14-FEB-2002.  
XX  
PF 08-AUG-2001; 2001WO-US024795.  
XX  
PR 08-AUG-2000; 2000US-00634341.  
XX  
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA (UTAH-) UNIV UTAH RES FOUND.  
XX  
PI Adderson E. Bohnsack J;  
XX  
DR WPI; 2002-257465/30.  
DR P-PSDB; AAU77626.  
XX  
PT New streptococcal matrix adhesion (Ema) polypeptides, useful as vaccines,  
PT particularly for treating or preventing infections by virulent forms of  
PT streptococci.  
XX  
PS Example 1; Page 147-150; 177pp; English.  
XX  
CC The invention relates to isolated streptococcal polypeptides, which  
CC comprise the Streptococcus agalactiae extracellular matrix adhesion (Ema)  
CC polypeptides EmaA, EmaB, EmaC, EmaD or EmaE, and their encoding nucleic  
CC acids. Also included are the a vaccine comprising an Ema protein, an anti  
CC -Ema antibody, a pharmaceutical composition comprising one or more Ema

CC proteins plus an anti Ema antibody, optionally in combination with at  
CC least one antibody to a protein selected from Sbp1 and Sbp2, Rib, Lmb,  
CC C5a-ase or C protein alpha antigen, an immortal cell line producing an  
CC anti-Ema monoclonal antibody, a vector comprising the Ema nucleic acid, a  
CC host cell transformed with the vector, a nucleic acid vaccine comprising  
CC the vector and Ema homologous proteins and their encoding nucleic acids,  
CC from additional bacterial species (S. pneumoniae, S. pyogenes,  
CC Enterococcus faecalis and Corynebacterium diptheriae). The streptococcal  
CC polypeptides are useful as vaccines, particularly for treating or  
CC preventing infections by virulent forms of streptococci, especially group  
CC B streptococci (GBS) the most common cause of serious bacterial disease  
CC in neonates. The present sequence is the S. agalactiae Sbp1 gene  
XX  
SQ Sequence 1509 BP; 534 A; 241 C; 298 G; 436 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1509; DB 6; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAAAGAAAATGATTCAATCGCTGTAGTGGAGTTTGTAGCAATTTGGTATGGCTGTA 60  
DB 1 ATGAAAAGAAAATGATTCAATCGCTGTAGTGGAGTTTGTAGCAATTTGGTATGGCTGTA 60  
QY 61 TCACCAGTTAGCCGATAGCTTTTCCGCTGAGACAGGCAATTCAGATTCAAGATACT 120  
DB 61 TCACCAGTTAGCCGATAGCTTTTCCGCTGAGACAGGCAATTCAGATTCAAGATACT 120  
QY 121 CAAAAGGCGCAACTCTATAAGCATATAAAGTTTTTGTAGCAGAAATAGATAATGCAAAAT 180  
DB 121 CAAAAGGCGCAACTCTATAAGCATATAAAGTTTTTGTAGCAGAAATAGATAATGCAAAAT 180  
QY 181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAAATTCCTCAAGGTAAGAAGCTGAG 240  
DB 181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAAATTCCTCAAGGTAAGAAGCTGAG 240  
QY 241 TATAAGCTTCAACTGATTTTAAATTCCTTTTATAGCACTACTTAATGAGGGGAGACA 300  
DB 241 TATAAGCTTCAACTGATTTTAAATTCCTTTTATAGCACTACTTAATGAGGGGAGACA 300  
QY 301 TATGTAACATAAAAAAGATACCTGCGTCAGCAATGAGATTGGCAGATGGGCTAAATCTATA 360  
DB 301 TATGTAACATAAAAAAGATACCTGCGTCAGCAATGAGATTGGCAGATGGGCTAAATCTATA 360  
QY 361 TCAGCTAATACTACACCAAGTTTCCACTGTTTACTGAGTCAAAATAATGATGAGGTT 420  
DB 361 TCAGCTAATACTACACCAAGTTTCCACTGTTTACTGAGTCAAAATAATGATGAGGTT 420  
QY 421 ATTAATGTTTCCCAATATGAGATATTTATGTTTCTAGCACTGTTTAAATGAGGCTGTA 480  
DB 421 ATTAATGTTTCCCAATATGAGATATTTATGTTTCTAGCACTGTTTAAATGAGGCTGTA 480  
QY 481 ATTATGTTTACATCTGTAATCCAAATGCTACTATTTCATGAAAGAAATGATGCGGACA 540  
DB 481 ATTATGTTTACATCTGTAATCCAAATGCTACTATTTCATGAAAGAAATGATGCGGACA 540  
QY 541 TGGGAGATGGTGGGAAAACCTGTAGATCAAAAAACGTTACTCGTTGGTGATACAGTC 600  
DB 541 TGGGAGATGGTGGGAAAACCTGTAGATCAAAAAACGTTACTCGTTGGTGATACAGTC 600  
QY 601 AAATATATCTATTACTTATAAGAAATGCAATTCATGTTGATGAGAAAGAGTGTATCAA 660  
DB 601 AAATATATCTATTACTTATAAGAAATGCAATTCATGTTGATGAGAAAGAGTGTATCAA 660  
QY 661 TATGTTATAAAGGATCTATGCGCATCTGCTTCTGTTAGTTGATTTGAACGAAAGGTTCTTAT 720  
DB 661 TATGTTATAAAGGATCTATGCGCATCTGCTTCTGTTAGTTGATTTGAACGAAAGGTTCTTAT 720  
QY 721 GAAGTAACTATTACTGATGGATCAGGGAATATTCAACTCTAACTCAAGGTTCCGAAAAA 780  
DB 721 GAAGTAACTATTACTGATGGATCAGGGAATATTCAACTCTAACTCAAGGTTCCGAAAAA 780  
QY 781 GCAACTGGGAAGTATACTCTTTAGAGGAAAATTAATTTTCAAGATTACTATTCCGTTGG 840  
DB 781 GCAACTGGGAAGTATACTCTTTAGAGGAAAATTAATTTTCAAGATTACTATTCCGTTGG 840

```
Db 781 GCNACTGGGAAGTATACCTGTTAGAGGAAATAATAATTTTCACGATTACTATTCCGTGG 840
QY 841 GCAGCTACCAATACCTCAACCGGAATACCTCAAAATGGAGCTAAATGATGACTTTTAT 900
Db 841 GCAGCTACCAATACCTCAACCGGAATACCTCAAAATGGAGCTAAATGATGACTTTTAT 900
QY 901 AAGGGAATAAATAACAATCAAGTCAGTCCTTATACAGGAGTATTAAAGAGTGGAGCTAAACCA 960
Db 901 AAGGGAATAAATAACAATCAAGTCAGTCCTTATACAGGAGTATTAAAGAGTGGAGCTAAACCA 960
QY 961 GGTTCAGCTGATTATACAGAAATAACAACATTCGACCATCAACCCCAATCTAGCAAT 1020
Db 961 GGTTCAGCTGATTATACAGAAATAACAACATTCGACCATCAACCCCAATCTAGCAAT 1020
QY 1021 GATGACCCAGGTCAAAAGTAACAGTCAGGAGTGGTCAAAATCTATATAAAAATTTGAT 1080
Db 1021 GATGACCCAGGTCAAAAGTAACAGTCAGGAGTGGTCAAAATCTATATAAAAATTTGAT 1080
QY 1081 GGTTCACAAAAGCTTCATTACAAAGTGTCTATATTTTAAAGAAATGCTACGGGTCAA 1140
Db 1081 GGTTCACAAAAGCTTCATTACAAAGTGTCTATATTTTAAAGAAATGCTACGGGTCAA 1140
QY 1141 TTTCTAACTTTAACGATACAAATAACGTTGAATGGGGCAGAGCTAATGCAACAGAA 1200
Db 1141 TTTCTAACTTTAACGATACAAATAACGTTGAATGGGGCAGAGCTAATGCAACAGAA 1200
QY 1201 TATACACAGGACAGATGATTAATTAACATTACAGGCTTGAAGAAGGTACATCTAT 1260
Db 1201 TATACACAGGACAGATGATTAATTAACATTACAGGCTTGAAGAAGGTACATCTAT 1260
QY 1261 CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTAGATACTCTCAGAAAGTTAT 1320
Db 1261 CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTAGATACTCTCAGAAAGTTAT 1320
QY 1321 TTAGGAGATGGAGCCACTGATACGATTAATTCAGATAACCTTTAGTTAACCCACTGTT 1380
Db 1321 TTAGGAGATGGAGCCACTGATACGATTAATTCAGATAACCTTTAGTTAACCCACTGTT 1380
QY 1381 GAAATAACAAAGTACTGAGTTGCCCTTCAACAGGTGGTATTGGTACAAACAATTTCTAC 1440
Db 1381 GAAATAACAAAGTACTGAGTTGCCCTTCAACAGGTGGTATTGGTACAAACAATTTCTAC 1440
QY 1441 ATTATAGTGCAATTTTATAGTAATAGGAGCAGGTATCGTGCTTGTGTCGTCTGCTGTTA 1500
Db 1441 ATTATAGTGCAATTTTATAGTAATAGGAGCAGGTATCGTGCTTGTGTCGTCTGCTGTTA 1500
QY 1501 CGTTCCTTAA 1509
Db 1501 CGTTCCTTAA 1509
```

## RESULT 3

AAAX20108  
ID AAAX20108 standard; DNA; 1983 BP.

AC AAAX20108;

XX 20-APR-1999 (first entry)

DE Enterococcus faecalis gene EF058.

XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;  
KW detection; attenuation; antigenic; ss.

OS Enterococcus faecalis.

XX WO980554-A2.

PN 12-NOV-1998.

XX 04-MAY-1998; 98WO-US008959.

XX 06-MAY-1997; 97US-0044031P.

PR 16-MAY-1997; 97US-004655P.  
PR 14-NOV-1997; 97US-0066009P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Kunsch CA, Choi GH, Bailey C, Hromockyj A;  
XX WPI; 1999-070095/06.  
XX P-PSDB; AAY00118.

XX New isolated Enterococcus faecalis polynucleotides - used to develop  
PT products for the detection of Enterococcus and for use in vaccines for  
PT prevention or attenuation of Enterococcus infection.

XX Claim 1; Page 133; 301pp; English.

XX The present sequence represents a gene isolated from Enterococcus  
CC faecalis. The present invention describes genes, proteins and antigenic  
CC polypeptides isolated from E. faecalis. The proteins can be used in  
CC vaccines for preventing or attenuating an infection caused by a member of  
CC the Enterococcus genus in an animal. The nucleotide sequences can be used  
CC Enterococcus antibodies in a sample. The nucleotide sequences can be used  
CC for detecting Enterococcus nucleic acids. Products from the present  
CC invention can also be used for screening compounds to identify agonists  
CC and antagonists of E. faecalis protein activity

XX Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;

Query Match 4.3%; Score 64.8; DB 2; Length 1983;

Best Local Similarity 52.1%; Pred. No. 0.00013;

Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCAACAGAAATATACACAGGAGCAGATGGTATAATTAATCATTACAGGC 1239

Db 1660 ACARAAGCTGAAGCAACTACTTTTACACAAACGGCTGATGGATTAGTTGATATCACAGG 1719

QY 1240 TTGAAGAAGGTACATATCTATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299

Db 1720 CTTAAATACGGTACCTATATTTAGAGAAACTGTAGCTCTCTGATGATTATGCTTTGTTA 1779

QY 1300 GATAACTCTCAGAAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359

Db 1780 ACAATTCGATTTGATTTTGTGTCATCAATCAATCATATGCACAA-----CAGAAAAC 1833

QY 1360 CTTTGTAGTTAACCCAACTGTTGAAAATAACAAAGTACTGAGTTGCCCTTCAACAGGTGCT 1419

Db 1834 CTAGTTTCACCAGAAAAAGTACCAACAAACAAAGGTACCTTACCTTCAACAGGTGCTC 1893

QY 1420 ATTGTTACACAATTTTCTACATTATAGTGTGCAATTTTAGTAATAGGAGCAGGTATCGTG 1479

Db 1894 AAAGGAATCTAGCTTACTTAGGAAGTGGCGAGTCTTGTCTACTTATTGCGAGGACTCTAC 1953

QY 1480 CTTGTTGCTCGTCGTCGTTTACGTTCTTAA 1509

Db 1954 TTTGCTAGACGTAGAAAAAGAAATGCTTAA 1983

## RESULT 4

ABN98093

ID ABN98093 standard; DNA; 1983 BP.

XX ABN98093;

AC 05-AUG-2002 (first entry)

XX E faecalis EF058 gene.

XX Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;  
KW gene; ds.

XX Enterococcus faecalis.

XX US2002045737-A1.

```
XX 18-APR-2002.
XX 04-MAY-1996; 98US-00071035.
XX 04-MAY-1998; 98US-00071035.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX WPI: 2002-425450/45.
XX P-PSDB; ABP43337.
XX
XX New genes and polypeptides from Enterococcus faecalis, useful as vaccines
XX for preventing, treating or attenuating an infection caused by a member
XX of the Enterococcus genus in an animal, particularly E. faecalis.
XX
XX Claim 1; Page 92; 255pp; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of polypeptides from Enterococcus faecalis. The proteins can be
XX used as vaccines for preventing or attenuating an infection caused by a
XX member of the Enterococcus genus in an animal, particularly E. faecalis.
XX The polynucleotide is also useful for preventing or treating E. faecalis
XX infection. The present sequence is a coding sequence of the invention
XX
XX Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;
XX
XX Query Match 4.3%; Score 64.8; DB 6; Length 1983;
XX Best Local Similarity 52.1%; Pred. NO. 0.00013;
XX Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;
XX
XX 1180 ACAGAAGCTAATGCAACAGAAATATACACAGGAGGAGATGATATATACCATTTACAGG 1239
XX 1660 ACAAAAGCTGAAGCAACTACTTTTACAACACCGCTGATGATGATGATATACAGG 1719
XX
XX 1240 TTGAAGAAGGTACATATCTAGTTGAGAAAAAGGCTCCCTTAGGTTTACAAATTTGTTA 1299
XX 1720 CTTAAATACGGTACCTATTATTATAGAAAGAACTGTAGCTCCTGATGATGATGTTGTTA 1779
XX
XX 1300 GATACTCTCAGAGGTTATTTTAGGAGTGGAGCCAGTACGATGACGACTAAATTCAGATAC 1359
XX 1780 ACAATTCGAGTTGAATTTGTGCTCAATGAACTATATGACAA-----CAGAAAC 1833
XX
XX 1360 CTTTGTAGTTAACCCAACTGTTGAAATAACAAAGGTACTGCTTGCCTTCAACAGTGGT 1419
XX 1834 CTAGTTTACACGAAAGTACCAACAAACAAAGGTACTTACCTTCAACAGTGGC 1893
XX
XX 1420 ATTGTACAAATTTTCTACATTTATAGTGCAATTTTAGTAAATAGGAGGATATCGTG 1479
XX 1894 AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTGCAAGGAGTCTAC 1953
XX
XX 1480 CTTGTTGCTCGTCGTTTACGTTCTTAA 1509
XX 1954 TTGCTAGACGTAGAAAGAAAGAAATGCTTAA 1983
XX
XX
XX RESULT 5
XX AC88057
XX ID AC88057 standard; DNA; 1983 BP.
XX AC AC88057;
XX
XX 07-JUL-2003 (first entry)
XX
XX E. faecalis novel gene #109.
XX
XX Gene; db; endocarditis; bacteraemia; urinary tract infection; UTI;
XX intraabdominal infection; soft tissue infection; neonatal sepsis;
XX vaccine.
XX
XX Enterococcus faecalis.
XX
```

US2003017495-A1.

23-JAN-2003.

29-JUL-2002; 2002US-00206576.

06-MAY-1997; 97US-0044031P.

16-MAY-1997; 97US-0046655P.

14-NOV-1997; 97US-0066009P.

04-MAY-1998; 98US-00071035.

(HUMA-) HUMAN GENOME SCI INC.

Choi GH, Bailey C, Hromockyj A, Kunsch CA;

WPI: 2003-416890/39.

P-PSDB; ABU88365.

New nucleic acid molecules and polypeptides from Enterococcus faecalis, useful as vaccines for preventing or attenuating an enterococcal infection in an animal, or for identifying Enterococcus faecalis in biological samples.

Claim 1; Page; 40pp; English.

The invention relates to a new isolated nucleic acid molecule comprising a polynucleotide isolated from Enterococcus faecalis appearing as AC897949-ACA88196 (or sequences complementary to them or 95% identical to them). Also included are the proteins encoded by the above nucleic acids, making a recombinant vector (comprising inserting the isolated nucleic acid molecule cited above into a vector), a host cell comprising the vector (used to produce the protein), an isolated antibody specific for the polypeptides, a hybridoma that produces the antibody, an isolated polypeptide antigen comprising an amino acid sequence of an Enterococcus faecalis epitope listed in the specification, a vaccine comprising one or more E. faecalis polypeptides (and a pharmaceutical diluent, carrier or excipient) where the polypeptide elicits protective antibodies in an animal to a member of the genus Enterococcus; preventing or attenuating an infection caused by a member of the genus Enterococcus in an animal comprising administering to the animal the polypeptide and detecting Enterococcus nucleic acids in a biological sample. The E. faecalis nucleic acid molecules and polypeptides are useful as vaccines for preventing or attenuating an enterococcal infection in an animal (e.g. endocarditis, bacteraemia, urinary tract infection (UTI), intraabdominal infection, soft tissue infection and neonatal sepsis). The polypeptides are also useful for detecting Enterococcus aureus in immunoassays, as epitope tags, as molecular weight markers, or for generating antibodies that specifically bind E. faecalis polypeptides. The nucleic acid molecules are also useful as probes for gene mapping, or for identifying E. faecalis in biological samples. The kit and methods are useful for detecting Enterococcus antibodies or nucleic acid molecules in a biological sample. The present sequence is a novel E. faecalis nucleic acid of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO at [seqdata.uspto.gov/sequence.html?docID=20030017495](http://seqdata.uspto.gov/sequence.html?docID=20030017495)

Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;

Query Match 4.3%; Score 64.8; DB 7; Length 1983; Best Local Similarity 52.1%; Pred. NO. 0.00013; Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAAGCTAATGCAACAGAAATATACACAGGAGGAGATGATATATACCATTTACAGG 1239

DB 1660 ACAAAAGCTGAAGCAACTACTTTTACAACACCGCTGATGATGATGATATACAGG 1719

QY 1240 TTGAAGAAGGTACATATCTAGTTGAGAAAAAGGCTCCCTTAGGTTTACAAATTTGTTA 1299

DB 1720 CTTAAATACGGTACCTATTATTATAGAAAGAACTGTAGCTCCTGATGATGATGTTGTTA 1779

QY 1300 GATACTCTCAGAGGTTATTTTAGGAGTGGAGCCAGTACGATGACGACTAAATTCAGATAC 1359

Db	1780	ACAAATCGGATTGAATTTGGTCAATGAACAAATCATATGGCAAA-----CAGAAAC	1833
QY	1360	CTTTTAGTTAAACCAACTGTTGAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTGGT	1419
Db	1834	CTAGTTTACCAGAAAGTACCAACAACAAAGGTACCTTACCTTTCAACAGGTGGC	1893
QY	1420	ATTGGTACAAATTTCTACATATATAGTGCAATTTTAGTAAATAGGAGCAGGTATCGTG	1479
Db	1894	AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTATTCAGGAGTCTAC	1953
QY	1480	CTTGTTCCTGCTGCTGTTTACGTTCTTAA	1509
Db	1954	TTTGCTAGACGTAGAAAAGAAATGCTTAA	1983
RESULT 6			
ID	ABX61663	standard; DNA; 1983 BP.	
XX	AC	ABX61663;	
XX	DT	26-FEB-2003 (first entry)	
DE	Enterococcus faecalis EF040	polynucleotide #109.	
XX	EF040;	gene; ds; immunostimulant; antibacterial; gene mapping.	
XX	OS	Enterococcus faecalis.	
XX	PN	US6448043-B1.	
XX	PD	10-SEP-2002.	
XX	PF	04-MAY-1998; 98US-00071035.	
XX	PR	06-MAY-1997; 97US-0044031P.	
PR	16-MAY-1997;	97US-0046655P.	
PR	14-NOV-1997;	97US-0066009P.	
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Choi GH, Bailey C, Hromockyj A, Kunsch CA;	
XX	WPI;	2003-089120/08.	
DR	P-PSDB;	ABU13616.	
XX	New EF040 polypeptides and polynucleotides from Enterococcus faecalis, useful for generating an immune response against E. faecalis and other Enterococcus species, and as vaccines against other bacterial genera.		
XX	PS	Example 1; Col 113-116; 146pp; English.	
XX	The invention relates to polynucleotide fragments of a gene from Enterococcus faecalis, EF040, and the polypeptides encoded by them. The polypeptides are useful in detecting E. faecalis, as epitope tags, as molecular weight markers on SDS-PAGE gels or for molecular sieve gel filtration columns, in generating antibodies that specifically bind to the E. faecalis polypeptides, in generating an immune response against E. faecalis and other Enterococcus species and as vaccines against other bacterial genera. The polynucleotides are useful as probes for gene mapping and for identifying E. faecalis in biological samples. Sequences CC ABX61555-ABX61802 represent EF040 polynucleotides of the invention. Note: CC The sequence data for this patent can also be obtained from USPTO at CC seqdata.uspto.gov/sequence.html		
XX	Sequence	1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;	
Query Match	4.3%;	Score 64.8; DB 7; Length 1983;	
Best Local Similarity	52.1%;	Pred. No. 0.00013;	
Matches 172;	Conservative	0; Mismatches 152; Indels 6; Gaps 1;	

QY	1180	ACAGAAGCTAATGCAACAGCAATATACACAGGAGCAGATGGTATAATTACCATTA	1239
Db	1660	ACAAAGCTGAAGCAACTACTTTTACAAACCGCTGATGGATTAGTTGATATCACAGGG	1719
QY	1240	TTGAAAGAGGTACATACACTATCTAGTTTGAGAAAAGGCTCCCTTAGGTTACAATTTGTTA	1299
Db	1720	CTTAAATACGGTACCTATTATTTAGAGAAACTGTAGCTCCTGATGATTATGTTCTGTTA	1779
QY	1300	GATPACTCTCAGAAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC	1359
Db	1780	ACAATCGGATTTGAATTTTGGTCAATGAACATCATATGGCACAA-----CAGAAAAC	1833
QY	1360	CTTTTAGTTAAACCAACTGTTGAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTGGT	1419
Db	1834	CTAGTTTACCAGAAAAGTACCAACAAAGGTACCTTACCTTCAACAGGTGGC	1893
QY	1420	ATTGGTACAAATTTCTACATATATAGTTGCAATTTTAGTAAATAGGAGCAGGTATCGTG	1479
Db	1894	AAAGGAATCTACGTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTATTCAGGAGTCTAC	1953
QY	1480	CTTGTTCCTGCTGCTGTTTACGTTCTTAA	1509
Db	1954	TTTGCTAGACGTAGAAAAGAAATGCTTAA	1983
RESULT 7			
ID	ABK11591	standard; DNA; 2199 BP.	
XX	AC	ABK11591;	
XX	DT	05-JUN-2002 (first entry)	
DE	E. faecalis	genomic region containing an EmaC/D homologue.	
XX	EX	Extracellular matrix adhesion; EmaC/D; ds; group B streptococcus; GBS;	
KW	DNA vaccine; Spbi; Spb2; Rib; Lmb; C5a-ase; C protein alpha antigen;		
XX	OS	neonatal bacterial infection.	
XX	PN	WO200212294-A2.	
XX	PD	14-FEB-2002.	
XX	PF	08-AUG-2001; 2001WO-US024795.	
XX	PR	08-AUG-2000; 2000US-00634341.	
XX	PA	(SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.	
XX	PA	(UTAH ) UNIV UTAH RES FOUND.	
XX	PI	Adderson E, Bohnsack J;	
XX	WPI;	2002-257465/30.	
XX	New streptococcal matrix adhesion (Ema) polypeptides, useful as vaccines, particularly for treating or preventing infections by virulent forms of streptococci.		
XX	Example 3;	Page 167-169; 177pp; English.	
XX	The invention relates to isolated streptococcal polypeptides, which comprise the Streptococcus agalactiae extracellular matrix adhesion (Ema) polypeptides EmaA, EmaB, EmaC, EmaD or EmaE, and their encoding nucleic acids. Also included are the a vaccine comprising an Ema protein, an anti-Ema antibody, a pharmaceutical composition comprising one or more Ema proteins plus an anti Ema antibody, optionally in combination with at least one antibody to a protein selected from Spbi and Spb2, Rib, Lmb, C5a-ase or C protein alpha antigen, an immortal cell line producing an anti-Ema monoclonal antibody, a vector comprising the Ema nucleic acid, a host cell transformed with the vector, a nucleic acid vaccine comprising the vector and Ema homologous proteins and their encoding nucleic acids		









ABQ39490;  
12-JUL-2002 (first entry)  
Oligonucleotide for detecting cytosine methylation SEQ ID NO 26081.  
Human, cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
drug; side effect; cancer; central nervous system; cardiovascular;  
gastrointestinal; respiratory system; single nucleotide polymorphism;  
SNP; cell differentiation; ds.  
Homo sapiens.  
WO200218632-A2.  
07-MAR-2002.  
01-SEP-2001; 2001WO-EP010074.  
01-SEP-2000; 2000DE-01043826.  
05-SEP-2000; 2000DE-01044543.  
(EPIG-) EPIGENOMICS AG.  
Olek A, Piepenbrock C, Berlin K, Guetig D;  
WPI; 2002-371829/40.  
Determining the degree of cytosine methylation in genomic DNA, useful for  
diagnosis and prognosis, comprises selective hybridization of amplicons  
from chemically treated DNA.  
Claim 12; 56pp + Sequence Listing; 56pp; German.  
This invention describes a novel method for determining the degree of  
methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
genomic sample of DNA. The sample is treated chemically to convert  
cytosine (C) but not methylated C, to uracil, then part of the genomic  
DNA that contains the target C is amplified to form a labeled amplicon.  
The amplicon is hybridised to two classes, each with at least one member,  
of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
degree of hybridisation to both classes is determined from the label on  
the amplicon. From the ratio of labels hybridised to the two classes of  
oligonucleotides, the degree of methylation is calculated. The method is used:  
(i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
and of a wide range of diseases, e.g. cancer, disorders of the central  
nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
particularly by detecting mutations or single nucleotide polymorphisms  
(SNP's); and (ii) for differentiation of cell or tissue types and for  
investigating cell differentiation. The method allows the methylation  
status of many C residues to be determined simultaneously. ABQ13410-  
ABQ54121 represent genomic DNA sequences used to illustrate the method  
for determining the degree of cytosine methylation described in the  
disclosure of the invention  
Sequence 969 BP; 177 A; 54 C; 93 G; 645 T; 0 U; 0 Other;  
Query Match 3.8%; Score 56.6; DB 6; Length 969;  
Best Local Similarity 42.5%; Pred. No. 0.0075;  
Matches 360; Conservative 0; Mismatches 484; Indels 3; Gaps 1;  
90 TGAGACAGGGAACAATACAGTTCAAGTACTCAAAAAGGCGCAACCTATAAGCATATAA 149  
849 TAAACGAATAAATAACGAATAAATACTAATAAAGCAATAAATAAATAAATAA 790  
150 AGTTTTTCATGCAGTAAGTAAATGCAATGATCTGATTCGAATAAAGAGCTTC 209  
789 TACTATAAAGCAATAAATAAATACGAATAAATACTAATAAAGCAATAAATAACGA 730  
210 TTATTTTAATTCCTCAAGGTAAGAGCTGAGTATAAGCTTCAACTGATTTTATCTCT 269  
729 TAAATACTAATAAAGCAATAAATAAATACGAATAAATAAATACTAATAAAGCAATAA 670

QY 270 TTTTACGACAACTACTAATGGAGGGAGACATATGTAACTAAAAAGATACATCCGTCAGC 329  
DB 669 AACGAATAAATACTAATAAAGCAATAAATAAAGCAATAAATAAATACTAATAAAGCA 610  
QY 330 AAATGAGATTGGACATGGGCTAAATCTATATCAGCTAATACTACACCACTTTCCACTGT 389  
DB 609 TAAAAATACGAATAAATAAATACTAATAAAGCAATAAATAAATACTAATAAAGCAATA 550  
QY 390 TACTGAGTCAAAATAAGTGGTACTGAGGGTTATTAATGTTTCCCAATATGATATTATTA 449  
DB 549 AACGAATAAATAAATAAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAT 490  
QY 450 TGTTCCTAGCCTGTTAATAATGAGCTGTAATATGTTTACATCTGTAATCAATGC 509  
DB 489 TAACGAAACGAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAAATA 430  
QY 510 TACTATTCAATAAAGCAATACTGATCGACATGGGAGATGGTGGTGAATAAATACTGAGA 569  
DB 429 AAATACTAATAAAGCAATACTGATCGAATAAATAAATAAATAAATAAATAAATAA 370  
QY 570 TCAAAAAACGCTACTCGGTGGTGATACAGTCAAAATATCTATTATTAAGAAATGCAGT 629  
DB 369 AAATAAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAAATAAATA 310  
QY 630 CAATTATCATGGTACAGAAAAGTGTATCAATATGTTTAAAGGATCTATGCCATCTGC 689  
DB 309 GAATACGAATACGAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAA 250  
QY 690 TTCTGCTAGTTGTAACGAAGGCTCTTATGAAGTAACTATTACTGATCGATCAGGAA 749  
DB 249 TACTAATAAAGCAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAA 190  
QY 750 T---ATTACAACTCTAAGTTCGGAAGGCAAGCACTGGGAGGATTAACCTGTTAGA 806  
DB 189 TAAAAATACTAATAAAGCAATAAATAAATAAATAAATAAATAAATAAATAAATAA 130  
QY 807 GGAATAAATAAATTCAGGATTAATTCCTGGGCGCTACCAATATCTCCCAACCGGAA 866  
DB 129 TAAAAATAAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAAATAA 70  
QY 867 TACTCAAAATGAGCTAATGATGATCTTTTATAAGGGAATAAATAAATAAATAAATA 926  
DB 69 AACGAACGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 10  
QY 927 TTATACA 933  
DB 9 TAATAAA 3  
RESULT 13  
ABQ39491  
ID ABQ39491 standard; DNA; 969 BP.  
XX ABQ39491;  
XX  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26082.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200218632-A2.  
XX  
XX  
PD 07-MAR-2002.  
XX  
XX 01-SEP-2001; 2001WO-EP010074.  
PF  
XX  
XX 01-SEP-2000; 2000DE-01043826.  
PR





CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 2408 BP; 724 A; 504 C; 437 G; 743 T; 0 U; 0 Other;
QY	Query Match 3.7%; Score 55.4; DB 4; Length 2408;
DB	Best Local Similarity 49.9%; Pred. No. 0.017;
QY	Matches 195; Conservative 0; Mismatches 191; Indels 5; Gaps 2;
DB	20 AATCGCTGTTAGTGGCGAGTTTAGCAATTTGGTATGGCTGTATCAC--CAGTTACCCCGAT 77
QY	963 ATTCACTTTAGCTGTAGCTTTAGCTTTAGCTTTAGCTTTGGCGTCATAACACAGATACAGATAC 1022
DB	78 AGCTTTGGCGGTGAGACAGGGGCAATATACAGTTCAAGTACTCAAAAAGGCGCAACCTA 137
QY	1023 AGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1082
DB	138 TAAAGCATATAAAGTTTTTGGATCGAGAAATAGATATGCAATGTATCTGTTCGAATAA 197
QY	1083 AGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1142
DB	198 AGATGAGCTTCTTTATTTAATTCCTCAAGTAAAGAGCTGAGTATAAAGCTTCAACTGA 257
QY	1143 AGATACAGATACAGATACAGATAC---AGATACAGATACAGATACAGATACAGATAC 1199
DB	258 TTTTAAATCTCTTTTACGACAACTACTAATGCGGGAGAACATATGTAACATAAAAAAGA 317
QY	1200 TACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGAT 1259
DB	318 TACTCGCTCAGCAAAATGAGATTCGACATGGGCTAAATCTATATCACTACAGTAACTAC 377
QY	1260 TACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGAT 1319
DB	378 AGTTTCCACTGTTACTCAGTCAAAATATGAT 408
QY	1320 TACAGATACAGATACAGATACAGATACAGAT 1350
DB	
RESULT 15	
ABL56203/c	
ID	ABL56203 standard; DNA; 32392 BP.
XX	
AC	ABL56203;
XX	
DT	01-JUL-2002 (first entry)
XX	
DE	AmEPV genome fragment#5.
XX	
KW	AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
KW	genetic deficiency disorder; ds.
XX	
OS	Amsacta moorei entomopoxvirus.
XX	
FN	WO200212526-A2.
XX	
PD	14-FEB-2002.
XX	
XX	10-AUG-2001; 2001WO-US025287.
PF	
XX	10-AUG-2000; 2000US-0224479P.
PR	
PR	14-SEP-2000; 2000US-00662254.
XX	
XX	(UYFL ) UNIV FLORIDA.
PA	
XX	Moyer RW, Li Y, Bawden AL;
PI	
XX	WPI; 2002-227161/28.
DR	
XX	Novel recombinant entomopox virus vector useful for delivering
PT	polynucleotide encoding protein to vertebrate cell, comprises
PT	polynucleotide encoding protein operably linked with heterologous
PT	promoter sequence.
XX	

PS	Disclosure; Page 226-242; 326pp; English.
XX	
CC	The invention relates to a recombinant entomopox virus (EPV) vector,
CC	comprising a polynucleotide encoding a protein operably linked with a
CC	heterologous promoter sequence. The invention also concerns methods for
CC	providing gene therapy for genetic deficiency disorders. Vectors of the
CC	invention are useful for delivering a polynucleotide encoding a protein
CC	to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC	The vector is introduced into the vertebrate cell by infection in a viral
CC	particle, or by transfection, transduction, or injection either in vitro
CC	or in vivo. The vector is useful for the delivery and expression of
CC	biologically useful proteins in gene therapy protocols, and for
CC	delivering large DNA segments for engineering of vertebrate cells.
CC	Polynucleotides of the invention have applications in techniques such as
CC	their use as insertion sites for foreign genes of interest, hybridisation
CC	probes, for chromosome and gene mapping, in PCR technologies, and in the
CC	production of sense or antisense nucleic acids. Vectors of the invention
CC	provide for stable integration and expression of heterologous DNA in host
CC	cells, and are adapted for accepting large heterologous polynucleotide
CC	inserts which can be delivered in an infected or transformed cell and
CC	expressed in a stable fraction. The current sequence represents a
CC	fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC	(AmEPV)
XX	
SQ	Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 U; 0 Other;
QY	Query Match 3.7%; Score 55.2; DB 6; Length 32392;
DB	Best Local Similarity 44.0%; Pred. No. 0.034;
QY	Matches 329; Conservative 0; Mismatches 413; Indels 6; Gaps 2;
DB	479 TAAATATGGTTTACATCTCTGTAATCTCAATGCTACTATTTCATGAAAAGATACATCGATCG 538
QY	1602 TAAATAATAAATGGATATATTAATTTGCGCTCTAAACTATCATATTTTTCAGATCATAT 1543
DB	539 CATGGGAGATGGTGGTGGAAAACTGTAGATCAAAAAACGCTCGGTGGTGATACAG 598
QY	1542 ACATATATGTTGATGATGAAGAAAATTAGAGATTTTAAATAAATAATAATCTCGTTTAA 1483
DB	599 TCAAAATATACATTAATTAAGAAATGAGTCAATATTCATCGGTACAGAAAAAGTGTATC 658
QY	1482 TTAATAATATATATTA---ACAATAATCTAAAAAATAATACTACAATAATAGATAAT 1426
DB	659 AATATGTTTATAAGGATACTGCCATCTGCTTCGTAGTTGATTGAAACGAAGGCTCT 718
QY	1425 ATAATGAAATAGAAAAAATAAATTAATTAATGAATTAAGTGTGTGATATAG---TG 1369
DB	719 ATGAAATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 778
QY	1368 ATTCGACAGATCTAACTAATAAATAAAGCAATATATATTTCAATCAGTATATAGTAAAA 1309
DB	779 AAGCACTGGGAAGTATTAACCTGTTAGAGGAAAAATAAATTTTCAGGATTAATTCCT 838
QY	1308 CAGAAGAGGGAGTAAATTTCTGTTATTAACGTGATATCTGATAATAAATAAATAAATGTT 1249
DB	839 GGGCAGCTACCAATCTCCAAACGGAAATACCTCAAAATGGAGCTAATGATGACTTTTTT 898
QY	1248 TAGCAGAGATATATCTTTTAAAAAGAGGATGTATAAAGAGATTTATAATATAGATTTAT 1189
DB	899 ATAAGGGAATAAATAACAATCAGATCACTTATACAGGAGTATTAAGAGTGGAGCTAAAC 958
QY	1188 ATGAGAAAAAATAATATGTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1129
DB	959 CAGGTTCACTGATTTTACCAGAAAAATACAAACATTCGCGACCATCAACCCCAATACCTAG 1018
QY	1128 TATCTTTTGTGATTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1069
DB	1019 ATGATGACCCAGGTCAAAAAGTAACAGTCAGGGATGGTCAAAATTAATAAATAAATAAAT 1078
QY	1068 TTATTAACGATGTATTAAATTTGAAAGCAAAATTCGAAAGAAAAAGAGCTATTTCAAAT 1009
DB	1079 ATGGTTCCCAAAAGCTTCATTAACAGGTCCTATATTTCTTTTAAAGATGCTACGGGTC 1138
QY	1008 CTATATATGAAAAAAGTTCAATTAGAAAAATATATTTTATTTTATTTTGAATTTAAAAAG 949
DB	



1261 CAGATCCAGATCCAGATACAGATACATATACAGATACAGATACAGATACATTTTACAGATA 1202  
185 CTGATTGGAATAAGATGGAGCTTCTTATTATTCCTCAAGTAAAGAGCTGAGTATA 244  
1201 CAGATACAGATACAGATACAGATACAGATACATTTTACACATACGGATACAGATACATATA 1142  
245 AGCTTCA--ACTGATTTTAACTCTTTTACGACACTACTAATCGAGGGAGACAT 301  
1141 CAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATA 1082  
302 ATGTAATAAAGATACCTGCGTCAGCAAAATGAGATTTCGACATCGGGCTAATCTATAT 361  
1081 CAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATA 1022  
362 CAGCTAATACATACACAGTTTCCACTGTTTACGTAGTCAATATGATGCTAGGTATA 421  
1021 CAGATACAGATACAGATACATATACAGATACAGATACAGATACAGATACAGATACAGATA 962  
422 TTAATGTTTCCCAATATGGA 441  
961 CAGATACAGTGGGAATGGA 942

RESULT 18  
ABQ75107/c  
ID ABQ75107 standard; CDNA; 4985 BP.  
XX AC ABQ75107;  
XX DT 01-NOV-2002 (first entry)  
XX DE Anopheles gambiae odourant receptor 2 genomic DNA SEQ ID NO:10.  
XX KW Anopheles gambiae; mosquito; olfactory gene; arrestin 1; pest control;  
XX KW odourant receptor; olfaction; gene; ds.  
XX OS Anopheles gambiae.  
XX PN WO200259274-A2.  
XX XX 01-AUG-2002.  
XX XX 28-JAN-2002; 2002WO-US002549.  
XX XX 26-JAN-2001; 2001US-0264649P.  
XX PR 24-JAN-2002; 2002US-00056405.  
XX XX (UYVA-) UNIV VANDERBILT.  
XX XX Zwiebel LJ;  
XX XX WPI; 2002-627421/67.  
XX DR P-PSDB; ABP52835.  
XX XX New mosquito olfaction polypeptides and polynucleotides, useful for  
PT mosquito management, i.e. controlling the pest and disease vectors, or  
PT for identifying pest control agents.  
XX PS Disclosure; Fig 4a; 96pp; English.

The present invention describes a purified Anopheles gambiae olfaction polypeptide comprising a 383, 394, 380, 411, 412, 391, 157 or 401 residue amino acid sequence (see ABP52833 to ABP52840) (S1), a conservatively modified amino acid sequence of them, or a sequence of (S1) with at least 20 consecutive residues. Also described: (1) an isolated polynucleotide comprising: (a) a nucleotide sequence encoding the purified Anopheles gambiae olfaction polypeptide; or (b) a nucleotide sequence that hybridises under stringent conditions to a hybridisation probe comprising a 1964, 1239, 1142, 1236, 1194, 1176, 474 or 1206 nucleotide sequence (see ABQ75102 to ABQ75105 and ABQ75110 to ABQ75113) (S2), or its complement; and (2) a method for identifying an agent that binds to mosquito olfaction molecules comprising: (a) providing an isolated

mosquito olfaction molecule; (b) contacting a test agent with the isolated mosquito olfaction molecule; and (c) detecting specific binding of the test agent to the isolated mosquito olfaction molecule, where the presence of specific binding identifies the test agent as a mosquito olfaction-binding compound. The mosquito olfaction molecules are useful for mosquito management, i.e. controlling this pest and disease vector. A method from the present invention of screening for substances that modulate arrestin-odourant receptor interaction is useful for identifying pest control agents. The present sequence represents Anopheles gambiae odourant receptor 2 genomic DNA from the present invention. N.B. The features given in figure 4a are tentative and do not directly encode SEQ ID NO:6 (ABP52835)

Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 U; 0 Other;

Query Match 3.5%; Score 53; DB 6; Length 4985;  
Best Local Similarity 44.9%; Pred. No. 0.069;  
Matches 241; Conservative 0; Mismatches 295; Indels 1; Gaps 1;  
QY 170 ATAATGCAAAATGTATCTGATTGGAATAAAGATGGAGCTTCTTATTTTCTCCTCAAGGTA 229  
DB 2854 ATAATAATATTTAAAAATAATATTAATAATAATAATAATAATAATAATAATAATA 2795  
QY 230 AAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATCTCTTTTACGACACTACTAATG 289  
DB 2794 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2735  
QY 290 GAGGGAGAACATATGTAATCTACTAAAAAGATGCTGCGTCAGCAAAATGAGATTTCGACATGGG 349  
DB 2734 TGATATTAATGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2675  
QY 350 CTAAATCTATATCA-GCTAATACTACACCCAGTTTCCACTGTTTCTGCTGACTCAAAATGAT 408  
DB 2674 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2615  
QY 409 GGTACTGAGGTATTAAATGTTTCCCAATATGATATTAATTTCTVAGCACTGTTAAT 468  
DB 2614 AATAATAATAATCATTAATTAATAATAATAATAATAATAATAATAATAATAATAATA 2555  
QY 469 AATGGAGCTGTAATTTATGTTTACATCTCTTAATCTCAAAATGCTACTATTTCATGAAAGAA 528  
DB 2554 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2495  
QY 529 ACTGATGGACATGGGGAGATGGTGGGAAACCTGATGATCAAAACCGTACTCGGT 588  
DB 2494 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2435  
QY 589 GGTGATACAGTCAAAATATATCTTATTAAGATGCGAGTCAATATCATGGGTACAGAA 648  
DB 2434 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2375  
QY 649 AAAGTGTATCAATATGTTTATAAAGGATGATGCGCATCTGCTTCTGTAGTTGTTG 705  
DB 2374 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2318

RESULT 19  
ACF79720/c  
ID ACF79720 standard; DNA; 4985 BP.  
XX AC ACF79720;  
XX DT 15-JAN-2004 (first entry)  
XX DE Mosquito odourant receptor 2 genomic DNA.  
XX KW Mosquito; odourant receptor 2; olfaction; insecticide; antimalarial; gene;  
XX KW ss.  
XX OS Anopheles gambiae.  
XX XX  
XX FH Key Location/Qualifiers  
XX FT exon 1..702

FT                   /\*tag= d  
FT                   /number= 1  
FT                   1. .485  
FT                   /\*tag= a  
FT                   295. .300  
FT                   /\*tag= b  
FT                   486. .3561  
FT                   /\*tag= c  
FT                   /product= "Odorant receptor 2"  
FT                   /notes= "contains introns"  
FT                   703. .798  
FT                   /\*tag= e  
FT                   /number= 1  
FT                   799. .882  
FT                   /\*tag= f  
FT                   /number= 2  
FT                   883. .978  
FT                   /\*tag= g  
FT                   /number= 2  
FT                   979. .1404  
FT                   /\*tag= h  
FT                   /number= 3  
FT                   1405. .1601  
FT                   /\*tag= i  
FT                   /number= 3  
FT                   1602. .1702  
FT                   /\*tag= j  
FT                   /number= 4  
FT                   1703. .1783  
FT                   /\*tag= k  
FT                   /number= 4  
FT                   1784. .1886  
FT                   /\*tag= l  
FT                   /number= 5  
FT                   1887. .1993  
FT                   /\*tag= m  
FT                   /number= 5  
FT                   1994. .2106  
FT                   /\*tag= n  
FT                   /number= 6  
FT                   2107. .3459  
FT                   /\*tag= o  
FT                   /number= 6  
FT                   3460. .4985  
FT                   /\*tag= p  
FT                   3562. .4985  
FT                   /\*tag= q  
FT                   WO2003076590-A2.  
PN  
XX  
XX  
PD  
XX  
PP  
XX  
PR  
XX  
PA  
XX  
PI  
XX  
DR  
XX  
PT  
XX  
PT  
XX  
PT  
XX  
PS  
XX  
CC  
XX  
CC  
XX  
CC  
XX  
CC  
XX

5'UTR  
TATA\_signal  
CDS  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon  
3'UTR  
WO2003076590-A2.  
18-SEP-2003.  
10-MAR-2003; 2003WO-US007174.  
08-MAR-2002; 2002US-00094240.  
(UYVA-) UNIV VANDERBILT.  
Zwiebel LJ;  
WPI; 2003-722331/68.  
New mosquito arrestin 1 and 2 genes and polypeptides, useful for identifying mosquito olfaction molecule binding compounds which reduce the ability of mosquitoes to locate sources of bloodmeal, e.g. humans.  
Disclosure; Fig 4a; 101pp; English.  
The present sequence is that of genomic DNA encoding a novel mosquito olfaction molecule, odorant receptor 2. A cDNA sequence for odorant receptor 2 is given in ACF97919. The invention provides 9 novel mosquito polypeptides and the nucleic acids encoding them. These are odorant receptor molecules 1-7 and arrestins 1 and 2. The odorant receptors

CC function in a ligand-induced signal transduction pathway for the  
CC activation of mosquito olfaction. Arrestin functions to inhibit the  
CC activated signal transduction cascade. Thus, the odorant receptors act as  
CC an 'on' switch, and arrestin as an 'off' switch for the odorant detection  
CC system of the mosquito. Methods are provided for identifying compounds  
CC that interfere with the operation of the mosquito olfactory system,  
CC particularly compounds that modulate arrestin 2 activity. These are  
CC useful for the control of mosquitoes, particularly by reducing their  
CC ability to locate sources of bloodmeal  
XX  
SQ Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 U; 0 Other;  
Query Match           3.5%; Score 53; DB 9; Length 4985;  
Best Local Similarity 44.9%; Pred. No. 0.069;  
Matches 241; Conservative 0; Mismatches 295; Indels 1; Gaps 1;  
Qy   170 ATAATGCAATGATCTGATTTCGAATAAGATGGAGCTTCTTATTATTCCTCAAGGTA 229  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2854 ATAATAATATTAAAAATAATAATTATTAATAATAATAATAATAATAATAATAATA 2795  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Qy   230 AAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTACGCAACTACTAATG 289  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2794 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2735  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Qy   290 GAGGAGAACATATGTAACTAAAGATACCTGCGTCAGCAAAATGAGATTGCGACATGGG 349  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2734 TGATAATAATGTATAATAATGAATAACAATAATAATAATAATAATAATAATAATAAT 2675  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Qy   350 CTAATCTATATCA-GCTAATACTACACCAGTTTCCACTGTTACTGAGTCAATAATGAT 408  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2674 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2615  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Qy   409 GGTACTGAGGTTTATTAATGTTTCCCAATATGGAATATTATTATGTTTCTAGCACTGTTAAT 468  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2614 AATAATAATAATCATATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 2555  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Qy   469 AATGAGCTGTAATTATGTTTACATCTGTAACCTCAATGCTACTATTCATGAAAGAAAT 528  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2554 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2495  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Qy   529 ACTGATCGGACATGGGAGATGGTGGTGGRAAACTGTAGATCAAAAACGCTACTCGGTT 588  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2494 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2435  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Qy   589 GGTGATACAGTCAAATATATCTATTATTAAGAAGATGCAGTCAATTATCATGGTACAGAA 648  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2434 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2375  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Qy   649 AAAGTGATCAATATGTTTAAAGGATCACTATGCCATCTGCTTCTGTAGTTGATTG 705  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
Db   2374 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2318  
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||   ||  
RESULT 20  
AAA70152  
ID   AAA70152 standard; DNA; 6033 BP.  
XX  
XX  
AC   AAA70152;  
XX  
XX  
DT   07-NOV-2000 (first entry)  
XX  
DE  
XX  
XX  
KW   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:285.  
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.  
XX  
OS   Plasmodium falciparum.  
XX  
XX  
PN   WO200025728-A2.  
XX  
PD   11-MAY-2000.  
XX  
PF   05-NOV-1999; 99WO-US026796.

XX PR 05-NOV-1998; 98US-0107131P.  
XX PA (HOFF/) HOFFMAN S.  
XX PA (CARU/) CARUCCI D.  
XX PA (GARD/) GARDNER M.  
XX PA (VENT/) VENTER J C.  
XX PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX DR WPI; 2000-365347/31.  
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,  
XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
XX PT diagnosis of P.falciparum infection.  
XX PS Disclosure; Page 493-495; 577pp; English.  
XX CC The present invention describes proteins and their fragments (I) encoded  
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
XX CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
XX CC vaccines against P. falciparum infection comprising (I) or (II). (I) and  
XX CC (II) are useful for the development of vaccines against P. falciparum  
XX CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
XX CC immunogens comprising the sequences of (I), are useful in the detection  
XX CC of infection with P. falciparum. Furthermore, (I) (especially when they  
XX CC are rfinos or secreted or membrane proteins) can aid the identification  
XX CC of drugs to treat or prevent P. falciparum infection, or they can be used  
XX CC to identify drug resistance in P. falciparum. Sequencing of the  
XX CC Plasmodium chromosome 2 and the subsequent identification of proteins  
XX CC encoded by it will help to expand our understanding of parasite biology,  
XX CC a process hampered by the complexity of the parasitic lifecycle, and  
XX CC provide new targets for vaccine and drug development. Parasite resistance  
XX CC to drugs and mosquito resistance to insecticides have led to a resurgence  
XX CC of malaria in many parts of the world, and there is a pressing need for  
XX CC vaccines and new drugs. AAAY00078 to AAA70287 and AAB18144 to AAB18352  
XX CC represent nucleotide and protein sequences given in the present  
XX CC invention, but which are not specifically mentioned within the  
XX CC specification  
XX SQ Sequence 6033 BP; 3019 A; 437 C; 707 G; 1870 T; 0 U; 0 Other;  
Query Match 3.5%; Score 53; DB 3; Length 6033;  
Best Local Similarity 43.6%; Pred. No. 0.072;  
Matches 236; Conservative 0; Mismatches 305; Indels 0; Gaps 0;  
390 TACTGACTCAAAATGATGCTGCTAGTGGTATTAATGTTTCCCAATATGATATTATTA 449  
978 TATAGAATGTAATAATAAATAATGATGATGATAATAATGATGATAATAATAATAA 1037  
450 TGTTCCTAGCAGCTTTAATAATGAGCTGTAATTATGTTTACATCTGTAACCTCAAATGC 509  
1038 TAATGATGATAATAATAATAATAATAATGATGATAATAATAATGATGATAATAATA 1097  
510 TACTATTTCATGAAAGAAATCTGATCGGACATGGGAGATGGTGGGAAAAAAGCTGTAGA 569  
1098 TAATAATAATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATA 1157  
570 TCAAAAACCTACTCGTGGTGGTACAGTCAATATACATCTATTACTTATAAGATCGAGT 629  
1158 TTATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1217  
630 CAATTATCATGGTACAGAAAGATGATCAATATGTTTATAAGAGATCATGTCATCTGC 689  
1218 TATGTTAAATATACAAAATATATTATACAAAACATTTTAAAGAGATAAAAAAATATTA 1277  
690 TTCGTGATGATTTGAACGAGGCTTATGAAGTAACATTTACTGATGATCAGGAA 749  
1278 CACAAATAATAAAAAAGAAATATTTTTCCTCACTAATGGAACATTTGTATATGTATAA 1337  
750 TATTACAACTCTCACTCAAGCTTCGGAAGAAAGCACTGGGAAGATATAACCTGTTAGAGGA 809  
1338 AAAAAAATCTTTGATTAAATAAATGAAGAAAGAAATAATAATAATAATAATAATAATA 1397

QY 810 AAATAATAATTTCCAGATTACTATTCCGTGGCGCAGCTACCAATACTCCAAACCGGAAATAC 869  
DB 1398 CGATAAAATTAATTAAGTTTAATAAATAATAATATTATGTTATTAACAACAATAGTGAACG 1457  
QY 870 TCAAAATGAGCTAATGATGACTTTTTTATAAGGGAATAAATACAATCACAGTCACCTTA 929  
DB 1458 TTGCAAAATTAACGGCATATGTTATTTTGTGGTATATTAGGAATAAGTAATATATTTA 1517  
QY 930 T 930  
DB 1518 T 1518  
RESULT 21  
ABN67916  
ID ABN67916 standard; DNA; 2703 BP.  
XX AC ABN67916;  
XX DT 01-JUL-2002 (first entry)  
XX DE Streptococcus polynucleotide SEQ ID NO 3745.  
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX OS Streptococcus agalactiae.  
XX PN WO200234771-A2.  
XX PD 02-MAY-2002.  
XX PF 29-OCT-2001; 2001WO-GB004789.  
XX PR 27-OCT-2000; 2000GB-00026333.  
XX PR 24-NOV-2000; 2000GB-00028727.  
XX PR 07-MAR-2001; 2001GB-00005640.  
XX PA (CHIR-) CHIRON SPA.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;  
XX PI Tettelin H;  
XX DR WPI: 2002-352536/38.  
XX DR P-PSDB; ABP27285.  
XX PT New Streptococcus protein for the treatment or prevention of infection or  
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for  
XX PT detecting a compound that binds to the protein.  
XX PS Claim 7; Page 3529; 4525pp; English.  
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
XX CC the specification. The proteins have antibacterial and antiinflammatory  
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
XX CC antibodies that bind (I) are used in the manufacture of medicaments for  
XX CC the treatment or prevention of infection or disease caused by  
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX CC biological sample. (I) is used to determine whether a compound binds to  
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX CC used as a vaccine or diagnostic composition. The disease caused by  
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be  
XX CC used in gene therapy. Antibodies to (I) are used for affinity  
XX CC chromatography, immunoassays, and distinguishing/identifying  
XX CC Streptococcus proteins









```

RESULT 26
AAD47004
ID AAD47004 standard; DNA; 3996 BP.
XX
AC AAD47004;
XX
DT 27-JAN-2003 (first entry)
DE Plasmodium falciparum BBP-5 DNA.
XX
KW Band 3 polypeptide; malarial infection; drug resistance; vaccine;
XX protozoicide; gene therapy; BBP-5; gene; ds.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..3996
FT /*tag= a
FT /product= "BBP-5 protein"
XX
PN WO200270542-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-US006415.
XX
PR 02-MAR-2001; 2001US-0272930P.
XX
PA (SELI-) ST ELIZABETH'S MEDICAL CENT INC.
XX
PI Chishti AH, Oh SS, Liu D, Goel V;
XX
WPI; 2002-759814/82.
DR
DR P-PSDB; AAE29358.
XX
XX New isolated Band 3 polypeptide which selectively binds to merozoite
PT surface protein-1, useful for the prevention and treatment of malarial
PT infection.
XX
XX Claim 45; Page 135-137; 163pp; English.
XX
CC The invention relates to an isolated Band 3 polypeptide that comprises
CC any of 4 20 residue amino acid sequences, or their fragments that bind to
CC an MSP-1 polypeptide or a polypeptide with any of 8 291-1331 base pair
CC sequences. The methods and compositions of the present invention are
CC useful for the prevention and treatment of malarial infection. The
CC present invention develops new and more improved methods based upon
CC inhibiting the particular interactions between the malarial parasite and
CC a cognate molecule present in the host and subsequently minimising
CC harmful side effects and drug resistance that may be due to non-specific
CC therapeutic approaches. The invention is useful in gene therapy. The
CC present invention also provides a vaccine for malaria. The present
CC sequence is Plasmodium falciparum BBP-5 DNA
XX
SQ Sequence 3996 BP; 1776 A; 334 C; 636 G; 1250 T; 0 U; 0 Other;
Query Match 3.3%; Score 50; DB 6; Length 3996;
Best Local Similarity 47.5%; Pred. No. 0.3;
Matches 149; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 366 TAATACTACACAGTTTCCACTGTTACTGAGTCAATAATAGATGGTACTGAGGTTATTAA 425
DB |||||
DB 1881 TAATAATAATGATGAGAAATTTGGTAAATAATAATACCAATAATAATAATAATAATAA 1940
QY |||||
QY 426 TGTTCCTCCATATGGATATTATATGTTCTAGCACTGTTAATAATGGAGCTGTTATAT 485
DB |||||
DB 1941 TAATAATAATATGTTGTAACAATAATAATAATAACAAATAATAATAAGGACAATAATAATAA 2000
QY |||||
QY 486 GGTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAGAAATCTGATGCGACATGGGG 545
DB |||||
DB 2001 TGATGGAATGGTAGTAGTAATAATAATAATAATATGATGATGATGACGAAGAGAGGA 2060
|||

QY 546 AGATGGTGGTGGAAAAAAGCTAGATCAAAAAACGTAAGTCTGGTTGGTGATACAGTCAAAATA 605
|||
DB 2061 TGATGAAGATGATAACAATAATAATAATGATGATGATGATGATGATGATGATGATGATGAT 2120
|||
QY 606 TACTATTACTTATAAGAATGCGATGCAATATATCATGCTGGTACAGAAAAGTGTATCAATATGT 665
|||
DB 2121 GGAAGATAATGATGAAGATAACGATGAGTATAATAATAATAGTATGATGATGATGATGATGAT 2180
|||
QY 666 TATAAAGGATACCTA 679
|||
DB 2181 AGAAAAAGATAGTA 2194
|||

RESULT 27
AAX20109
ID AAX20109 standard; DNA; 1687 BP.
XX
AC AAX20109;
XX
DT 20-APR-1999 (first entry)
DE Enterococcus faecalis EF058 gene fragment.
XX
DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX detection; attenuation; antigenic; ss.
XX
KW Enterococcus faecalis.
XX
OS Enterococcus faecalis.
XX
PN WO9850554-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US008959.
XX
PR 06-MAY-1997; 97US-0044031P.
XX
PR 16-MAY-1997; 97US-0046655P.
XX
PR 14-NOV-1997; 97US-0066009P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Choi GH, Bailey C, Hromockyj A;
XX
WPI; 1999-070095/06.
DR
DR P-PSDB; AAY00119.
XX
XX New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines for
PT prevention or attenuation of Enterococcus infection.
XX
PS Claim 1; Page 134; 301pp; English.
XX
XX The present sequence encodes an antigenic polypeptide fragment isolated
CC from Enterococcus faecalis. The present invention describes genes,
CC proteins and antigenic polypeptides isolated from E. faecalis. The
CC proteins can be used in vaccines for preventing or attenuating an
CC infection caused by a member of the Enterococcus genus in an animal. They
CC can also be used for detecting Enterococcus antibodies in a sample. The
CC nucleotide sequences can be used for detecting Enterococcus nucleic
CC acids. Products from the present invention can also be used for screening
CC compounds to identify agonists and antagonists of E. faecalis protein
CC activity
XX
SQ Sequence 1687 BP; 632 A; 284 C; 360 G; 411 T; 0 U; 0 Other;
Query Match 3.3%; Score 49.8; DB 2; Length 1687;
Best Local Similarity 62.4%; Pred. No. 0.28;
Matches 78; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1180 ACAGAAGCTAATGCAACAGAAATATACAACAGAGCAGATGGTATTAATACATTACAGGC 1239
DB |||||
DB 1466 ACAAAAGCTGAAGCAACTACTTTTACAAACACGGCTGATGATGATGATGATGATGATGATGAT 1525
|||
QY 1240 TTGAAGAAGGTACTACTACTACTAGTTGAGAAAAGGCTCCCTTAGGTACAAATTTGTTA 1299
|||

```

Db	1526	CTTAATACGGTACCTATTATTATTAGAGAACTGTAGCTCCTGATGATTATGCTTGTTA	1585
Qy	1300	GATAA 1304	
Db	1586	ACAAA 1590	
RESULT 28			
ABN98094			
ID	ABN98094	standard; DNA; 1687 BP.	
AC	ABN98094;		
XX	05-AUG-2002	(first entry)	
DT	XX		
DE	E faecalis EF058	gene fragment.	
XX	Enterococcus;	vaccine; gastrointestinal disease; diagnosis; antibiotic;	
KW	Gene; ds.		
OS	Enterococcus faecalis.		
XX	US2002045737-A1.		
PN	18-APR-2002.		
XX	04-MAY-1998;	98US-00071035.	
PF	04-MAY-1998;	98US-00071035.	
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA	Choi GH, Bailey C, Hromockyj A, Kunsch CA;		
XX	WPI; 2002-425450/45.		
DR	P-PSDB; ABP43338.		
XX	New genes and polypeptides from Enterococcus faecalis, useful as vaccines for preventing, treating or attenuating an infection caused by a member of the Enterococcus genus in an animal, particularly E. faecalis.		
PS	Claim 1; Page 93-94; 255pp; English.		
XX	The present invention provides the protein and coding sequences of a number of polypeptides from Enterococcus faecalis. The proteins can be used as vaccines for preventing or attenuating an infection caused by a member of the Enterococcus genus in an animal, particularly E. faecalis. The polynucleotide is also useful for preventing or treating E. faecalis infection. The present sequence is a coding sequence of the invention		
XX	Sequence 1687 BP; 632 A; 284 C; 360 G; 411 T; 0 U; 0 Other;		
SQ	Query Match	3.3%; Score 49.8; DB 6; Length 1687;	
	Best Local Similarity	62.4%; Pred. No. 0.28;	
	Matches	78; Conservative 0; Mismatches 47; Indels 0; Gaps 0;	
Qy	1180	ACAGAAGCTTAATGCAACAGATATACACAGGACAGATGGTATTAATACCATTAACAGGC	1239
Db	1466	ACAAAGCTGAGCACTACTTTTACACACGGCTGATGATTAGTTGATATCACAGG	1525
Qy	1240	TTGAAGAAGGTACATACCTACTTACTGTGAGAAAAAGGTCCTTAGGTTACAAATTGTTA	1299
Db	1526	CTTAAATACGGTACCTATTATTATTAGAGAACTGTAGCTCCTGATGATTATGCTTGTTA	1585
Qy	1300	GATAA 1304	
Db	1586	ACAAA 1590	
RESULT 29			
ACA88058			
ID	ACA88058	standard; DNA; 1687 BP.	

XX	ACA88058;		
AC	07-JUL-2003	(first entry)	
DT	XX		
DE	E. faecalis novel gene #110.		
XX	Gene; ds; endocarditis; bacteraemia; urinary tract infection; UTI; intraabdominal infection; soft tissue infection; neonatal sepsis; vaccine.		
KW	Enterococcus faecalis.		
OS	US2003017495-A1.		
XX	23-JAN-2003.		
PF	29-JUL-2002; 2002US-00206576.		
XX	06-MAY-1997; 97US-0044031P.		
PR	16-MAY-1997; 97US-0046655P.		
PR	14-NOV-1997; 97US-0066009P.		
PR	04-MAY-1998; 98US-00071035.		
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	Choi GH, Bailey C, Hromockyj A, Kunsch CA;		
PI	WPI; 2003-416890/39.		
DR	P-PSDB; ABU88366.		
XX	New nucleic acid molecules and polypeptides from Enterococcus faecalis, useful as vaccines for preventing or attenuating an enterococcal infection in an animal, or for identifying Enterococcus faecalis in biological samples.		
PT	Claim 1; Page; 40pp; English.		
XX	The invention relates to a new isolated nucleic acid molecule comprising a polynucleotide isolated from Enterococcus faecalis appearing as ACA87949-ACA88196 (or sequences complementary to them or 95% identical to them). Also included are the proteins encoded by the above nucleic acids, making a recombinant vector (comprising inserting the isolated nucleic acid molecule cited above into a vector), a host cell comprising the vector (used to produce the protein), an isolated antibody specific for the polypeptides, a hybridoma that produces the antibody, an isolated polypeptide antigen comprising an amino acid sequence of an Enterococcus faecalis epitope listed in the specification, a vaccine comprising one or more E. faecalis polypeptides (and a pharmaceutical diluent, carrier or excipient) where the polypeptide elicits protective antibodies in an animal to a member of the genus Enterococcus; preventing or attenuating an infection caused by a member of the genus Enterococcus in an animal comprising administering to the animal the polypeptide and detecting Enterococcus nucleic acids in a biological sample. The E. faecalis nucleic acid molecules and polypeptides are useful as vaccines for preventing or attenuating an enterococcal infection in an animal (e.g. endocarditis, bacteraemia, urinary tract infection (UTI), intraabdominal infection, soft tissue infection and neonatal sepsis). The polypeptides are also useful for detecting Enterococcus aureus in immunoassays, as epitope tags, as molecular weight markers, or for generating antibodies that specifically bind E. faecalis polypeptides. The nucleic acid molecules are also useful as probes for gene mapping, or for identifying E. faecalis in biological samples. The kit and methods are useful for detecting Enterococcus antibodies or nucleic acid molecules in a biological sample. The present sequence is a novel E. faecalis nucleic acid of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO at seqdata.uspto.gov/sequence.html?docID=20030017495		
XX	Sequence 1687 BP; 632 A; 284 C; 360 G; 411 T; 0 U; 0 Other;		
SQ	Query Match	3.3%; Score 49.8; DB 7; Length 1687;	

	Best Local Similarity Matches 78; Conservative	62.4%; 0; Mismatches 47; Indels	Pred. No. 0.28; 0; Gaps	0; 0;
QY	1180 ACAGAAGCTTAATGCAACAGATATACACAGGACAGATGGTATAATTACCATTTACAGGC	1239		
Db	1466 ACAAAAGCTGAAGCAACTACTCTTTTACAAACAACGGCTGATGGATTAGTTGATATCACAGGG	1525		
QY	1240 TTGAAAGAAGGTACATCTACTCTAGTTTGAGAAAAAGGCTCCCTTAGGTTACAAATTTGTTA	1299		
Db	1526 CTTAAATACGGTACCTATTATTTTAAAGAAGAACTGTAGTCTCCTGATGATTATGTCCTTGTTA	1585		
QY	1300 GATAA 1304			
Db	1586 ACAAA 1590			
Search completed: July 26, 2004, 01:29:14				
Job time : 675 secs				

RESULT 30	
ABX61664	
ID	ABX61664 standard; DNA; 1687 BP.
XX	
XX	ABX61664;
XX	
XX	AC
XX	
XX	ABX61664;
XX	
XX	DT
XX	26-FEB-2003 (first entry)
XX	
XX	DE
XX	Enterococcus faecalis EF040 polynucleotide #110.
XX	
XX	EF040; gene; ds; immunostimulant; antibacterial; gene mapping.
XX	
XX	OS
XX	Enterococcus faecalis.
XX	
XX	FN
XX	US6448043-B1.
XX	
XX	PD
XX	10-SEP-2002.
XX	
XX	XX
XX	04-MAY-1998; 98US-00071035.
XX	
XX	PR
XX	06-MAY-1997; 97US-0044031P.
XX	
XX	PR
XX	16-MAY-1997; 97US-0046655P.
XX	
XX	PR
XX	14-NOV-1997; 97US-0066009P.
XX	
XX	PR
XX	14-NOV-1997; 97US-0066009P.
XX	
XX	PA
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	
XX	PI
XX	Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX	
XX	DR
XX	WPI; 2003-089120/08.
XX	
XX	DR
XX	P-PSDB; ABU13617.
XX	
XX	PT
XX	New EF040 polypeptides and polynucleotides from Enterococcus faecalis,
XX	useful for generating an immune response against E. faecalis and other
XX	Enterococcus species, and as vaccines against other bacterial genera.
XX	
XX	PS
XX	Example 1; Col 115-116; 146pp; English.
XX	
XX	CC
XX	The invention relates to polynucleotide fragments of a gene from
XX	Enterococcus faecalis, EF040, and the polypeptides encoded by them. The
XX	polypeptides are useful in detecting E. faecalis, as epitope tags, as
XX	molecular weight markers on SDS-PAGE gels or for molecular sieve gel
XX	filtration columns, in generating antibodies that specifically bind to
XX	the E. faecalis polypeptides, in generating an immune response against E.
XX	faecalis and other Enterococcus species and as vaccines against other
XX	bacterial genera. The polynucleotides are useful as probes for gene
XX	mapping and for identifying E. faecalis in biological samples. Sequences
XX	ABX61555-ABX61802 represent EF040 polynucleotides of the invention. Note:
XX	The sequence data for this patent can also be obtained from USPTO at
XX	seqdata.uspto.gov/sequence.html
XX	
XX	XX
XX	Sequence 1687 BP; 632 A; 284 C; 360 G; 411 T; 0 U; 0 Other;
XX	
XX	Query Match
XX	Best Local Similarity 3.3%; Score 49.8; DB 7; Length 1687;
XX	Matches 78; Conservative 62.4%; Pred. NO. 0.28;
XX	Mismatches 0; Mismatches 47; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 22:18:49 ; Search time 4180 Seconds  
(without alignments)  
10780.390 Million cell updates/sec

Title: US-10-009-254-1  
Perfect score: 1509  
Sequence: 1 atgaaaagaataattc.....gtcgtcgttcacgttcttaa 1509

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 90 summaries

Database : EST:

1: em\_estba:\*

2: em\_estbha:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estro:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_esthum:\*

18: em\_estinv:\*

19: em\_estpln:\*

20: em\_estvrt:\*

21: em\_estfun:\*

22: em\_estmam:\*

23: em\_estmus:\*

24: em\_estpro:\*

25: em\_estrod:\*

26: em\_estphg:\*

27: em\_estvrl:\*

28: gb\_est1:\*

29: gb\_est2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 4	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 2	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 3	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 4	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 5	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 6	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 7	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 8	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 9	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 10	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 11	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 12	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 13	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 14	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 15	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 16	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 17	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 18	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 19	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 20	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 21	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 22	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 23	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 24	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 25	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 26	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 27	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 28	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 29	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 30	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 31	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 32	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 33	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 34	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 35	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 36	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 37	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 38	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 39	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 40	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 41	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 42	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 43	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 44	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 45	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 46	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 47	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 48	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 49	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 50	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 51	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 52	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 53	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 54	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 55	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 56	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 57	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 58	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 59	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 60	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 61	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 62	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 63	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 64	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 65	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 66	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 67	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 68	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 69	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 70	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 71	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 72	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 73	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 74	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 75	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 76	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 77	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 78	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 79	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 80	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 81	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 82	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 83	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 84	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 85	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 86	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 87	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 88	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69.4	4.6	494	29	FR0048073
c 89	69	4.6	829	29	BX173672
	-----				
	76	5.0	843	28	B2643413
c 90	76	5.0	843	28	B2643413
	69.8	4.6	844	29	BX139987
	69				

C 78 53.6 3.6 500 28 B07882`  
 C 79 53.6 3.6 592 12 B07882`  
 C 80 53.6 3.6 650 28 A2573029 312PVG08  
 C 81 53.6 3.6 684 12 B07882`  
 C 82 53.6 3.6 828 12 B07882`  
 C 83 53.6 3.6 897 29 C07882`  
 C 84 53.6 3.6 987 29 C07882`  
 C 85 53.4 3.5 487 28 B07882`  
 C 86 53.4 3.5 594 12 B07882`  
 C 87 53.4 3.5 723 28 B07882`  
 C 88 53.4 3.5 723 29 C07882`  
 C 89 53 607 12 B07882`  
 C 90 53 742 9 A060168

# ALIGNMENTS

RESULT 1  
 BZ643413/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 843)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other GSs: OGAW57C  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..843  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZM0.7.1.5\_KB"  
 /clone\_lib="ZM 0.7.1.5\_KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN  
 Query Match 5.0%; Score 76; DB 28; Length 843;  
 Best Local Similarity 43.7%; Pred. No. 6.6e-07;  
 Matches 334; Conservative 0; Mismatches 430; Indels 0; Gaps 0;

QY 171 TAATGCAATGATCTGATTCGATTAAGATGAGCTTCTTATTTATTCCTCAAGTAA 230  
 DB 778 TACTGCTGCTGATCTGCTACTGCTACTGCTGCTACTGCTACTGCTACTGCTGC 719  
 QY 231 AGAAGCTGAGTAAAGCTCACTGATTTTAAATCTCTTTTACGACACTACTAATGG 290  
 DB 718 TACTGCTACTGCTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTGC 659  
 QY 291 AGGGAGACATATGTAACAAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 350  
 DB 658 TGCTACTACTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTAT 599

QY 351 TAAATCTATACAGTAAATACACACAGTTTCCACTGTTTACTGAGTCAATATATGATGG 410  
 DB 598 TGTACTCCAGCTACTACTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAC 539  
 QY 411 TACTGAGGTATTAAATGTTTCCCAATATGGAATATTTATTTTCTAGCACTGTTAATAA 470  
 DB 538 TGCTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAC 479  
 QY 471 TGGAGCTGTAATATGTTATCATCTGTAATCTCAAAATGCTACTATTATCATGAAGAATAC 530  
 DB 478 TGCTGCTGCAACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAC 419  
 QY 531 TGATGCGCATGGGAGATGCTGGAAATCTGTAGATCAAAAAAGTACTCGGTTGG 590  
 DB 418 TACTGCTACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359  
 QY 591 TGATACAGTCAAAATATATCTATTATTAAGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 650  
 DB 358 TACTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299  
 QY 651 AGTGTATCAATATGTTATTAAGGATATCTATGCCATCTGCTTCTGCTAGTTGATTTGAACGA 710  
 DB 298 AGCTACTACAGTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239  
 QY 711 AGGGTCTTATGAAGTAACTATTACTGATGAGTACGAGGAATATTACAATCTTAATCAAGG 770  
 DB 238 TACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 179  
 QY 771 TTCGAAAAAGCACTGGGAAGTATTAACCTGTTAGAGAAAAATTAATTTTACGATTTAC 830  
 DB 178 TGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 119  
 QY 831 TATTCGCTGGGAGCTACCAATCTCCAAACCGAAATCTCAAAATGAGCTGATTAATGATGA 890  
 DB 118 TAACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 59  
 QY 891 CTTTTTTTATAGGGAATTAATAACAATCACAGTCACTTTATACAG 934  
 DB 58 TACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15

RESULT 2  
 BXL39987  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Danio rerio (zebrafish)  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 844)  
 Humphray, S.J., Huckle, E. and Durham, J.L.  
 Direct Submission  
 Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome  
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humphray@sanger.ac.uk Unpublished  
 This sequence was generated from the T7 end of BAC 99E7. 99E7 is  
 part of the Daniokey BAC Library created by R. Plasterk and N.V.  
 Keygene. Further details:  
 http://www.sanger.ac.uk/projects/D\_rerio/  
 Location/Qualifiers  
 1..844  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-99E7"  
 /tissue\_type="Testis"  
 /note="vector pIndigoBAC-536"

BXL39987 844 bp DNA linear GSS 13-MAR-2003  
 Danio rerio genomic clone DKEY-99E7, genomic survey sequence.  
 BXL39987  
 BXL39987.1 GI:27971314  
 GSS.  
 Danio rerio (zebrafish)  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 844)  
 Humphray, S.J., Huckle, E. and Durham, J.L.  
 Direct Submission  
 Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome  
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humphray@sanger.ac.uk Unpublished  
 This sequence was generated from the T7 end of BAC 99E7. 99E7 is  
 part of the Daniokey BAC Library created by R. Plasterk and N.V.  
 Keygene. Further details:  
 http://www.sanger.ac.uk/projects/D\_rerio/  
 Location/Qualifiers  
 1..844  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-99E7"  
 /tissue\_type="Testis"  
 /note="vector pIndigoBAC-536"

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 source

ORIGIN



```
Query Match      4.6%; Score 69.8; DB 29; Length 844;
Best Local Similarity 46.3%; Pred. No. 1.8e-05;
Matches 230; Conservative 0; Mismatches 267; Indels 0; Gaps 0;

QY 137 ATAAAGCATATAAAGTTTGTGATGCAGAAATAGATAATCGAAATGATCTGATTCGAATA 196
    |||||
DB 348 ATATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 407

QY 197 AAGATGGAGCTTCTTTATTTTAAATTCCTCAAGGTAAGAGCTGAGTATATAAGCTTCAACTG 256
    |||||
DB 408 TTGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 467

QY 257 ATTTTAATTTCTTTTACGACAACTACTAATGAGGGGAGAACATATGTAACTTAAAAAAG 316
    |||||
DB 468 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 527

QY 317 ATACTGGCTGCAGCAATAGATTTGGACATGGCTGGCTAAATCTATATACGCTATATCTACAC 376
    |||||
DB 528 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 587

QY 377 CAGTTTCCACTGTTACTGAGTCAAAATGATGGTACTGAGGTTATTAATGTTTCCCAAT 436
    |||||
DB 588 ATGATGATAATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATA 647

QY 437 ATGGATATTTATGTTTCTAGCACTGTTAATAATGAGAGCTTAATAATGTTGATCACTG 496
    |||||
DB 648 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 707

QY 497 TAACTCCAAATCTACTATTTCAATGAAAGAACTGATCGGACATGGGGAGAGTGGTG 556
    |||||
DB 708 CTAATAAGACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 767

QY 557 GAAAACTGTAGATCAAAAAAGCTACTCGTTGGTGATACAGTCAAAATATATCTATTACTT 616
    |||||
DB 768 ATAATAAGTACTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 827

QY 617 ATAAGATGCAGTCAAT 633
    |||||
DB 828 ATAATAATGATGATAAT 844

RESULT 3
FR0048073
LOCUS      494 bp      DNA      linear      GSS 05-JAN-2001
DEFINITION Fugu rubripes GSS sequence, clone 263K15BD8, genomic survey
ACCESSION  AL444858.1  GI:12052694
VERSION     GSS; genome survey sequence.
KEYWORDS    Takifugu rubripes (Fugu rubripes)
SOURCE      Takifugu rubripes
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE   1 (bases 1 to 494)
            Direct Submission
            Author(s)      Clark, M.S.
            Title           Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource
            Journal         Centre Hinxton, Cambridge, CB10 1SB. UK Email:
            Vector           biohelp@hgmrc.ac.uk
            V type           phagemid
            PRIMER          KS
DESCR       One pass dye-terminator sequencing of BAC (pBelobACII) cloned.
            genomic sequence
            The BACs can be obtained from http://www.incyte.com.
FEATURES    source
            Location/Qualifiers
                1..494
                /organism="Takifugu rubripes"
                /mol_type="genomic DNA"

Query Match      4.6%; Score 69; DB 29; Length 829;
Best Local Similarity 45.2%; Pred. No. 2.7e-05;
Matches 252; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

/db_xref="taxon:31033"
/clone="263K15BD8"
/clone_lib="BAC 263K15"

ORIGIN
Query Match      4.6%; Score 69.4; DB 29; Length 494;
Best Local Similarity 48.6%; Pred. No. 2.1e-05;
Matches 190; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 236 CTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTACTAATGGAGGA 295
    |||||
DB 22 CTGGGCTTCTACTACTACTACTAGTCTGCTGCTACTACTACGACTACGACTACTGCTGCGT 81

QY 296 GAACATATGTAACTAAAAAGATAGTGGTGCAGCAATGAGATGCGCATGGGCTAAAT 355
    |||||
DB 82 CGACTACTGCGACGAGCTGCTACTACTACTGCGGCTGCTTCTACTACTGCTACTACTA 141

QY 356 CTATATCAGCTTAATACTACACAGTTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 415
    |||||
DB 142 CGAGACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 201

QY 416 AGTTTATTAATGTTTCCCAATATGATATTTATGTTTCTAGCACTGTTAATAATGAG 475
    |||||
DB 202 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 261

QY 476 CTGTAATTTATGTTTACTACTCTGTAATCTCAAAATGCTACTATTTCATGAAAGAAATCTGATG 535
    |||||
DB 262 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 321

QY 536 CGACATGGGAGAGTGGTGGGAAAAACTGTAGATCAAAAAAGCTACTCGGTTGGTGATA 595
    |||||
DB 322 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 381

QY 596 CAGTCAAAATATATCTATTACTTTATAAGATGC 626
    |||||
DB 382 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 412

RESULT 4
BX173672/c
LOCUS      829 bp      DNA      linear      GSS 13-MAR-2003
DEFINITION Danio rerio genomic clone DKEY-150M6, genomic survey sequence.
ACCESSION  BX173672.1  GI:28005377
VERSION     GSS.
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 829)
            Direct Submission
            Author(s)      Humphray, S.J., Huckle, E. and Durham, J.L.
            Title           Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Journal         Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            Vector           humquerry@sanger.ac.uk Unpublished
            Comment          This sequence was generated from the SP6 end of BAC 150M6. 150M6 is
            Comment          part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Comment          Keygene. Further details:
            Comment          http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES    source
            Location/Qualifiers
                1..829
                /organism="Danio rerio"
                /mol_type="genomic DNA"
                /db_xref="taxon:7955"
                /clone="DKEY-150M6"
                /tissue_type="Testis"
                /note="vector pIndigoBAC-536"

ORIGIN
Query Match      4.6%; Score 69; DB 29; Length 829;
Best Local Similarity 45.2%; Pred. No. 2.7e-05;
Matches 252; Conservative 0; Mismatches 305; Indels 0; Gaps 0;
```

```
QY 385 ACTGTTACTGAGTCAATATATGATGCTACTGAGGTTTATTAATGTTTCCCAATATGATAT 444
Db 645 AATAAATAAATAAGATAATAATAATAATGATGATGATGATGATGATGATGATGATGATGAT 586
QY 445 TATTATGTTTCTAGCACTGTTAATAATGAGCTGTAATATGTTTACATCTGTAATCCCA 504
Db 585 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 526
QY 505 AATGCTACTATTCAAGAAAGATACTGATGCGACATGGGAGATGGTGGTGGAAAAAAT 564
Db 525 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 466
QY 565 GTAGATCAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATATTTACTTTAAGAT 624
Db 465 GATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 406
QY 625 GCAGTCAAAATATCATCGTACAGAAAAAGTGATCAATATGTTAATAAGGATATGATGCA 684
Db 405 GATAAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 346
QY 685 TCTGCTTCTGATGTTGAACGAAGGCTCTTATGAAGTAACTATTTACTGATGATCA 744
Db 345 AATAAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 286
QY 745 GGAATAATTTACAACTCTAACTCAAGCTCGGAAAAAGCAACTGGGAAGTATTAACCTGTTA 804
Db 285 AATAAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 226
QY 805 GAGGAAAAATAAATTTTCAGATTACTATTTCGTTGGGCGACTACCAATACTCCAAACCGGA 864
Db 225 AATAAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 166
QY 865 AATACTCAAAATCGAGCTAATGATGACTTTTTTTTATAAGGATAAATAACAATCACAGTC 924
Db 165 AATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 106
QY 925 ACTTATACAGGAGTATT 941
Db 105 TAAATAGAAAATATT 89

RESULT 5
CNS02HAA4
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
139021 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL197365
GI:7835515
GSS; genome survey sequence.
Tetraodon nigroviridis
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Bernot,A., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
```

```
20359837
10899143
3 (bases 1 to 989)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
source
1..989
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="139021"
/clone_lib="G"
/note="Genoscope sequence ID : COAG139AH11LPL1-end : T7"

ORIGIN
Query Match 4.5%; Score 68; DB 29; Length 989;
Best Local Similarity 44.9%; Pred. No. 4.6e-05;
Matches 213; Conservative 7; Mismatches 254; Indels 0; Gaps 0;
QY 171 TAATCAAAATGATCTGATTCGAATAAAGATGGAGCTTCTTTATTTAAATCTCAAGTAA 230
Db 56 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 115
QY 231 AGAAGCTGAGTATAAGCTTCAACTGATTTTAAATTTCTCTTTTACGACAACTTAATGG 290
Db 116 TAATACAAAACACTACWACWACTACTACTACTACTACTACTACTACTACTACTACTAA 175
QY 291 AGGAGACATATGTAATAAAGATAGTCCGTCAGCAATAGATGGGACATGGC 350
Db 176 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 235
QY 351 TAAATCTATATCAGCTAATACTACACACCTTTCCACTCTTACTAGTCAAAATATGATGG 410
Db 236 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAA 295
QY 411 TACTGAGGTTAATGTTTCCCAATATGATTTATTTATTTTCTTCTAGCAGCTTTAA 470
Db 296 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 355
QY 471 TGGAGCTGTAATTTGTTTACATCTGTAATCTCAATGCTACTATTTCATGAAAAGATAC 530
Db 356 TACTGCTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 415
QY 531 TGATCGACATGGGAGATGGTGGGAAAACCTAGATCAAAAACGTAATCGGTGG 590
Db 416 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 475
QY 591 TGATCAGTCAATATATCTACTTACTTATAAGATCGAGTCAATATCATGGTAC 644
Db 476 TGGTCTCTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTCTCC 529

RESULT 6
AQ946120/c
LOCUS
DEFINITION
Sheared DNA-46J23-TR Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-46J23, genomic survey sequence.
AQ946120
ACCESSION
VERSION
GI:6769385
KEYWORDS
SOURCE
ORGANISM
Trypanosoma brucei
Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 641)
REFERENCE
AUTHORS
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
```

Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S.,  
 Donelson, J., Fraser, C. and Adams, M.  
 Determination of clone end sequences from Trypanosoma brucei GUTat  
 10.1 sheared DNA library  
 Unpublished (1999)  
 Other GSSs: Sheared DNA-46J23.TP  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org  
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
 DNA library constructed at TIGR. Clones will be available for  
 distribution through ATCC. Sheared DNA end sequences search page:  
<http://www.tigr.org/tldb/mdb/tbdb/>.  
 Seq primer: M13-Reverse  
 Class: shotgun.

#### FEATURES

source  
 1. .641  
 /organism="Trypanosoma brucei"  
 /mol\_type="genomic DNA"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone\_lib="Sheared DNA-46J23"  
 /note="Vector: pUC18; Site 1: SmaI; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically  
 sheared to give a tight size distribution (approx 2 kb).  
 The v + i method used for the library construction is  
 described in detail in Smith, H.O. and Venter, J.C.  
 (Making small insert libraries for whole genome shotgun  
 sequencing projects. In Genome Sequencing: A Practical  
 Approach, eds. M. Vaudin and B. Borell, Oxford University  
 Press, 1999)."

#### ORIGIN

Query Match 4.5%; Score 67.8; DB 28; Length 641;  
 Best Local Similarity 45.3%; Pred. No. 5e-05;  
 Matches 246; Conservative 0; Mismatches 297; Indels 0; Gaps 0;

QY 366 TAATACTACACAGTTTCCACTGTTTACTGAGTCAAAATATGATGCTACTGAGTTTATTA 425  
 |||||  
 Db 621 TAATAGTACTACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 562  
 |||||  
 QY 426 TGTTCCTCCAAATGATATATTATGTTCTTAGCACTGTTAATAATGAGCTGTAATTAT 485  
 |||||  
 Db 561 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 502  
 |||||  
 QY 486 GGTACATCTGTAACCTCAAAATGCTACTATTTCATGAAAGAACTACTGATGCGACATGGG 545  
 |||||  
 Db 501 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 442  
 |||||  
 QY 546 AGATGGTGGTGGAAAACCTGTAGATCAAAAACGCTACTCGGTGGTGTGATACAGTCAAAATA 605  
 |||||  
 Db 441 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 382  
 |||||  
 QY 606 TACTATTACTTATAGAATGCGAGTCAATTATCATCGGTACAGAAAAGTGTATCAATATGT 665  
 |||||  
 Db 381 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 322  
 |||||  
 QY 666 TATRAAGATACACTATGCCATCTGCTCTGTAGTCTGTTTGTACGAGGCTCTTATGAGT 725  
 |||||  
 Db 321 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 262  
 |||||  
 QY 726 AACTATTACTGATGAGTCAAGGAATATTACAACTCTAACTCAAGGTTCGGAAGAAAGCAAC 785  
 |||||  
 Db 261 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 202  
 |||||  
 QY 786 TGGGAAGTATACCTGTTAGAGGAAAAATAATAAATTTTCACGATTACTTCCGTGGGCGAC 845

Db 201 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 142  
 |||||  
 QY 846 TACCAATACTCCACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTTATAAGG 905  
 |||||  
 Db 141 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 82  
 |||||  
 QY 906 AAT 908  
 |||||  
 Db 81 AGT 79

#### RESULT 7

CG372306  
 LOCUS  
 DEFINITION  
 OG4BI08TC ZM.0.7.1.5 KB Zea mays genomic clone ZMMBma0811B15,  
 genomic survey sequence.

ACCESSION  
 CG372306

VERSION  
 CG372306.1 GI:34289573

KEYWORDS  
 GSS.

SOURCE  
 Zea mays

ORGANISM  
 Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE  
 1 (bases 1 to 619)

AUTHORS  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,

Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

Consortium for Maize Genomics

Unpublished (2002)

COMMENT  
 Contact: Cathy Whitelaw

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

FEATURES

Location/Qualifiers

1. .619

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBma0811B15"

/clone\_lib="ZM.0.7.1.5 KB"

/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

Query Match 4.4%; Score 67; DB 29; Length 619;

Best Local Similarity 45.5%; Pred. No. 7.6e-05;

Matches 238; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

QY 48 TGGTATGGCTGTATCACCAGTTACCGCGATAGCTTTTCCCGCTGAGACAGACAAATPAC 107

Db 53 TGCTACTGCTGCTACTACTACTCTGCTACTACTGCTACTACTACTACTACTACTACTAC 112

QY 108 AGTTCAAGATACTCAAAAGCGCAACCTATAAAGCATATAAAGTTTGTGATGCAAGAAAT 167

Db 113 TGCTACTACTACTACTACTACTACTCTCTACTACTACTACTACTACTACTACTACTAC 172

QY 168 AGATAATGCAATGTCATCTGATTCGAATAAAGATGGAGCTTCTTATTTTAAATTCCTCAAGG 227

Db 173 TACTACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTGCTACTACTACTGC 232

QY 228 TAAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTACGACAACTACTTAA 287

Db 233 TACTACTACTGCTACAACTACTCTCTGCTACTACTACTACTACTACTACTACTACTACTAC 292

QY 288 TGGAGGGAGAAATATGTAATAAAAAAGATACCTGCGTCAGCAAAATGAGATTGGACATG 347

Dd	281	TACTACTACTACTGCTACT
----	-----	---

QY 263 ATTCTCTTTTACGACAACTACTAATGGAGGAGAACATATGTAACTAAAAAGATACGTG 322  
Db 203 ATACTGGTTCTGGTACTGTGATGATCTGGTACTGATCTGGTACTGATCTGGTACTG 262  
QY 323 CGTGAGCAATGAGATTGGGACATGGCTGAAATCTATATACGCTAATACACTACACAGTTT 382  
Db 263 GTACTGATAGTGGTCTGGTACTGGTACTGATGATCTGGTCTGGTACTGATCTGGTACTG 322  
QY 383 CCACGTGTACTGAGTCAATATATGATGGTACTGAGGTATTAATGTTTCCCAATATGATG 442  
Db 323 GTACTGATAGTGGTCTGGTACTGGTCTGGTACTGATGATCTGGTACTGATCTGGTACTG 382  
QY 443 ATTATATGTTTCTAGCACCTGTTAATATGAGGAGCTGTAATATGTTTACATCTGTAAC 502  
Db 383 GTACTGATAGTGGTCTGGTACTGATGATCTGGTACTGATGATCTGGTACTGATCTGGTACTG 442  
QY 503 CAAATGCTACTATTATGAAAGAAATATCTGATGCGACATGGGAGATGGTGGTGAAGAAA 562  
Db 443 ATACTGGTACTGGTACTGGTACTGATGATCTGGTCTGGTACTGATGATCTGGTACTG 502  
QY 563 CTGTAGATCAAAAAAGTACTCGTTGGTGTGATACAGTCAAAATATATCTATTATAAGA 622  
Db 503 GTACTGATAGTGGTCTGGTACTGATGATCTGGTCTGGTACTGATGATCTGGTACTGAT 562  
QY 623 ATGCGATCAATATATCATGTGTACAGAAAAGTGTATCAATATGTTTATAAGGATATGTC 682  
Db 563 ATACTGGTACTGGTACTGATGATCTGGTACTGATGATCTGGTACTGATGATCTGGTACTG 622  
QY 683 CATCTGCTCTCTAGTGTGATTTGAAAGGAGGTCTTATGAACTAATCTATCTGAT 738  
Db 623 GTACTGGTACTGATCTGGTACTGATGATCTGGTACTGATGATCTGGTACTGAT 678

RESULT 10  
BZ780846  
LOCUS  
DEFINITION  
BZ780846  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
COMMENT

BZ780846 575 bp DNA linear GSS 14-MAR-2003  
ii20h11.g1 WGS-ShicolorF (DH5a methyl filtered) Sorghum bicolor  
genomic clone ii20h11, genomic survey sequence.  
BZ780846  
GSS.  
Sorghum bicolor (sorghum)  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 575)  
Rabinowicz, P.D., O'Shaughnessy, A.L., Balijs, V., Dedhia, N.,  
Katzenburger, E., King, L., Miller, B., Muller, S., Nascimento, L.,  
Zutavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.  
Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)  
Unpublished (2002)  
Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mcombie@cshl.org  
Plate: ii20 row: h column: 11  
Seq primer: -21M13UnivRev  
Class: shotgun  
High quality sequence stop: 575.  
Location/Qualifiers  
1. 575  
/organism="Sorghum bicolor"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4558"  
/clone="ii20h11"  
/lab\_host="DH5a"  
/clone\_lib="WGS-ShicolorF (DH5a methyl filtered)"  
/note="Site 1: Xba I; Site 2: Xba I; The vector was  
digested with XbaI and one nucleotide was added by fill in

FEATURES  
source

in the recessive 3' end. The genomic DNA was nebulized,  
end repaired, adaptor ligated and size fractionated using  
sephadex. The resulting fragments were between 0.8 and 3  
kb and were cloned into the vector (x/y reads in M13mp19,  
-b/g reads in pUC19). The same ligation was transformed  
into DH5a."

## ORIGIN

Query Match 4.4%; Score 66.2; DB 28; Length 575;  
Best Local Similarity 45.1%; Pred. No. 0.00012;  
Matches 245; Conservative 0; Mismatches 298; Indels 0; Gaps 0;

QY 137 ATAAGCAATATAAGTTTTCATGCGAGAAATAGATAATGCAAAATGCTATCTGATTCGAATA 196  
Db 3 ATA 62  
QY 197 AAGATGGAGCTTCTTATTTAATTCCTCAAGGTAAAGAGCTGAGTATATAAGCTTCAACTG 256  
Db 63 GTAGTAGTAGTAGTAATAATAGTAAATAATACTAATAATAATAGTAAATAATAATAGTA 122  
QY 257 ATTTTAATCTCTTTTACGACAACTACTAATGGAGGAGAACATATGTAACATAAAAAAG 316  
Db 123 ATAATAATAATAATGATACTAATAGTAAATAATAGGAGTAGTAATAATAGTAAATAATA 182  
QY 317 ATACTGCGTCAGCAAAATGAGATTCGGACATGGGCTAAATCTATATCAGCTAATACTAC 376  
Db 183 ATAATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 242  
QY 377 CAGTTTCCACGTGTACTGAGTCAATATGATGGTACTGAGGTTATTAATGTTTCCCAAT 436  
Db 243 ATAATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 302  
QY 437 ATGGATATTTATTTCTTAGCCTGTAAATAATGGAGCTGTAAATATGGTTACATCTG 496  
Db 303 ATAATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 362  
QY 497 TAACCTCAAATGCTACTATTTCATGAAAGAATACTGATGCGACATGGGAGATGGTGGT 556  
Db 363 ATAATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 422  
QY 557 GAAAACTGTAGATCAAAAAACGCTACGTTGGTGTATACAGTCAAAATATATCTTACTT 616  
Db 423 CTAATAGTAGTAATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 482  
QY 617 ATAAGAATGCACTCAATATCATGCTACAGAAAAAGTGTATCAATATGTTTAAAGGATA 676  
Db 483 GTAATAATAGTAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 542  
QY 677 CTA 679  
Db 543 CTA 545

## RESULT 11

CNS012TP 1101 bp DNA linear GSS 26-JUL-1999  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN08A01 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ALI02007  
ALI02007.1 GI:5613618  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

Db	313	AAAAAAAAAAAAAAAAAAAAAGGGCANGNNNGNCNCANGCCCTTACCCNGGAC	254
Qy	1110	TATATTTGTTTTAAAGAAATCTACGGTCAATTTCTAAACTTTAAAGATACAAATAACGT	1163
Db	253	CCNCNCACACACAAANAAGCGNGAGNNMNNAGGCCAACACACAGNNAACAAAGG	194
Qy	1170	TGAATGGGGCAGACAACCTAATGCAACAGAAATATACAACAGAGCA	1215
Db	193	AAAGAGAGGWAANAAGAAAAAANAAGAAWNGCGNNACRGGGGA	148
RESULT 12			
CNS0039G/c			
LOCUS			
DEFINITION			
Drosophila melanogaster genome survey sequence TET3 end of BAC #			
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit			
fly), genomic survey sequence.			
ACCESSION			
VERSION			
AL063921.1			
GI:4941778			
GSS.			
Drosophila melanogaster (fruit fly)			
Drosophila melanogaster			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE			
1 (bases 1 to 1101)			
AUTHORS			
Genoscope.			
Direct Submission			
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :			
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrs@genoscope.cns.fr			
- Web : www.genoscope.cns.fr)			
COMMENT			
Determination of this BAC-end sequence was carried out as part of a			
collaboration with the Berkeley Drosophila Genome Project (BDGP).			
The BDGP is constructing a physical map of the Drosophila			
melanogaster genome using these BACs. For further information			
please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila			
melanogaster BAC library was prepared by Kazutoyo Osoegawa and			
Aaron Mamoser in Pieter de Jong's laboratory in the Department of			
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,			
NY. The library is named RPCI-98 and was constructed by partial			
EcoRI digestion of Drosophila DNA provided by the BDGP from the			
isogenic strain y2; cn bw sp, the same strain used for the BDGP's			
p1 and EST libraries. A more detailed description of the library			
and how to order individual BAC clones the entire library, or			
filters for hybridization from the BACPAC Resource Center can be			
found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .			
LOCATION/Qualifiers			
source			
1. .1101			
/organism="Drosophila melanogaster"			
/mol_type="genomic DNA"			
/db_xref="taxon:7227"			
/clone="BACR08K10"			
/clone_lib="RPCI-98"			
/note="end : TET3"			
ORIGIN			
Query Match			
Best Local Similarity 4.3%; Score 65.2; DB 29; Length 1101;			
Matches 127; Conservative 259; Mismatches 271; Indels 5; Gaps 1;			
Qy	92	AGACAGGGACAAATTCACAGTACTCAAAAGCGCAACCTATAAGCATATAAG	151
Db	1059	DRWMDGAGTGWTTATWTTWWTTATWTTWTTWTTWTTWTTWTTWTTWTTWTTW	1000
Qy	152	TTTTTGATGCAAAATAGATAATGCAATGTATCTGATTCGAATAAAGATGGAGCTTCTT	211
Db	999	RDRGAGKRDRAATDADGAGRDRGKRKDKKDRKGDGDDKGGKKKKAATAKATKW	940
Qy	212	ATTTAATTTCTCAAGTAAGAAGCTGAGTATAAAGCTTCAACTGATTTAAATCTCTTT	271
Db	939	DDMDWDKDKWKGAKDKKADDDDDGAGDKDDGDKKADDDDDGDKDDDKDDDKAD	880
Qy	272	TTACGACAACTACTAATGAGGAGGAACAATATGTAACTAAAAAAGATACCTGCGTCAGCA	331

Determination of this BAC-end sequence was carried out as part of a			
collaboration with the European Drosophila Genome Project (EDGP) -			
<a href="http://www.edgp.ebi.ac.uk">http://www.edgp.ebi.ac.uk</a> - This Drosophila melanogaster BAC			
library (Dros BAC) was made by Alain Billaud at CEPH (Centre			
d'Etude du Polymorphisme Humain) with funding provided by a MRC			
project grant. The DNA was prepared from embryos by Alain Bucheton			
and Genevieve Payan. It has been constructed in the vector			
pBeloBAC11.			
Location/Qualifiers			
source			
1. .1101			
/organism="Drosophila melanogaster"			
/mol_type="genomic DNA"			
/db_xref="taxon:7227"			
/clone="BACN08A01"			
/clone_lib="DrosBAC"			
/plasmid="pBeloBAC11"			
/note="end : T7"			
ORIGIN			
Query Match			
Best Local Similarity 4.4%; Score 66.2; DB 29; Length 1101;			
Matches 269; Conservative 131; Mismatches 486; Indels 0; Gaps 0;			
Qy	330	AAATGAGATTGCGACATGGCTAAATCTATATACAGCTAATACACAGTTTCCACTGT	389
Db	1033	AWKWSWMMGWAARAGGVVMTGAWAAGHTGSWABKTTTITGTTKWKCKTWATSKWBT	974
Qy	390	TACTGAGTCAATAATCATGCTACTGAGGTATTAATGTTTCCCAATATGGAATATTATTA	449
Db	973	AWTTAAKWWAAMDAGNBTTCKAKWTTKMTTAWMTKAWTTKMTTAAATATKRTTBWBA	914
Qy	450	TGTTTCTAGCACCTGTTAAATATGAGCTGTAATATGTTTACATCTGTAACCTCAAAATGC	509
Db	913	TATGTTTGTWYTABTATWTGCHKRAWTTTMMWATASMYAAATVTAATRAKATRAGWA	854
Qy	510	TACTATTTCATGAAGAATACTGATGCGACATGGGAGATGGTGGTGGAAAACTGTAGA	569
Db	853	TMCWAKAAAAKAWBATANANGKGTCTGKBNWMTTITTKHTTYTASCTWTWTATWYAKA	794
Qy	570	TCAAAAAACCTGCTCGGTGCTGATACAGTCAATATACATTAATTAAGAATCCAGT	629
Db	793	MGWTTWTCNAATNGATTMTACWYWGATATWCTTACWATTTTGTGKAWTTAAGAH	734
Qy	630	CAATTATCATGGTACAGAAAAGTGTATCAATATGTTTATAAGGATACATGCACTGTC	689
Db	733	YAAWATATWMCKGAAGTAACTHRTATATAKTATWTTAAGAAATWMMWKGAMKGTITW	674
Qy	690	TTCTGTAGTTGATTGTTGAACGAGGCTTATGAAGTAACTATTACTGATGATCAGGAA	749
Db	673	TTCKWMMGMAATTTHTGTTCTMTGKWMGTGMMTTHCATTTTTTKYTTGTTTWA	614
Qy	750	TATTACAACTTAACCTCAAGGTTTCGGAAGAAAGCAACTGGGAAGTATAACCTGTTAGAGGA	809
Db	613	KGMTWTVAMAGKAWTTGWAAGAAAAAATAAATAAATAAATAAATAAATAAATAA	554
Qy	810	AAATAATAATTTTCAGATTACTATTCGCTGGGAGCTACCAATCTCCAAACCGGAATAC	869
Db	553	AAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	494
Qy	870	TCAAAATGGAGCTAATCATGACTTTTTTTATAGGGAATTAATACATCACTCAGTCACTTA	929
Db	493	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	434
Qy	930	TACAGGAGTATTAAGAGTGGAGCTAAACCCAGGTTTCAGCTGATTTCAGAAAAATACA	989
Db	433	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	374
Qy	990	CATTGCGACCATCAACCCCAATCTAGCAATGATGACCCAGGTCATAAAGTAACAGTGAG	1049
Db	373	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	314
Qy	1050	GGATGGTCAATTAATAAATAAATAATGATGGTTCCACAAAAGCTTCATTACAAAGTGC	1109







Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

#### FEATURES

source  
1. .480  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMWBMa0111J17"  
/clone\_lib="ZM.0.7.1.5\_KB"  
/notes="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

#### ORIGIN

Query Match 4.2%; Score 62.8; DB 28; Length 480;  
Best Local Similarity 48.8%; Pred. No. 0.0007;  
Matches 200; Conservative 0; Mismatches 207; Indels 3; Gaps 1;  
QY 130 GCAACCTATAAGCATATAAGTTTGGATGCGAATAATAGATAATGCAATGATCTGAT 189  
Db 17 GCTACTACTGCTACTACTGCTACTGCTGCTGCACTGCTACTGCTACTGCTACTGCT 76  
QY 190 TCGAATAAGATGAGCTTCTTATTTAAATTCCTCAAG---GTAAAGAGCTGAGTATAA 246  
Db 77 ACTACTACTACTGCTGCTACTACTACTGCTGCTGCTGCTACTACTACTACTGCTACT 136  
QY 247 GCTTCAACTGATTTAAATCTCTTTTTCGACAACTACTAAATGGAGGAGAACATATGA 306  
Db 137 ACTACTACTACTGCTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 196  
QY 307 ACTAAAAAGATAGCTGGTCAGCAATGAGATTGCGACATGGGCTAAATCTATATCAGCT 366  
Db 197 ACTGCTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256  
QY 367 AATACTACACCACTTTCCAGTGTACTGAGTCAATTAATGATGGTACTGAGGTATTAAT 426  
Db 257 ACTACTGCTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTACT 316  
QY 427 GTTTCCTCAATGATGATTTATTTATGTTTCTAGCACTGTTAAATAGGAGCTGTAATATG 486  
Db 317 GCTACTACTGCTACTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 376  
QY 487 GTTACATCTGTAATCTCAATGCTACTATTTCATGAAAGAAATGATGTC 536  
Db 377 ACTGCTGCTACTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTACT 426

#### RESULT 15

BZ423698  
LOCUS  
DEFINITION  
513 bp DNA linear GSS 10-DEC-2002  
id52c10.g1 WGS-SbicolorF (DH5a methyl filtered) Sorghum bicolor  
genomic clone id52c10 5', genomic survey sequence.  
BZ423698  
VERSION  
BZ423698.1 GI:26373196  
KEYWORDS  
GSS.  
SOURCE  
Sorghum bicolor (sorghum)  
ORGANISM  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 513)  
REFERENCE  
Rabinowitz,P.D., O'Shaughnessy,A.L., Ballija,V., Dedhia,N.,  
Katzburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,  
Zutavern,T., Palmer,L., McCombie,W.R. and Martienssen,R.A.  
Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)  
Unpublished (2002)  
CONTACT: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mcombie@cshl.org  
Plate: id52 row: c column: 10  
Seq primer: -21M13UnivRev  
Class: shotgun  
High quality sequence stop: 513.

#### FEATURES

source  
1. .513  
Location/Qualifiers

/organism="Sorghum bicolor"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4558"  
/clone="id52c10"  
/lab\_host="DH5a"  
/clone\_lib="WGS-SbicolorF (DH5a methyl filtered)"  
/note="Site 1: Xba I; Site 2: Xba I; The vector was  
digested with XbaI and one nucleotide was added by fill in  
in the recessive 3' end. The genomic DNA was nebulized,  
end repaired, adaptor ligated and size fractionated using  
sephadex. The resulting fragments were between 0.8 and 3  
kb and were cloned into the vector (.X/Y reads in M13mp19,  
.b/g reads in pUC19). The same ligation was transformed  
into DH5a."

#### ORIGIN

Query Match 4.1%; Score 62.6; DB 28; Length 513;  
Best Local Similarity 46.3%; Pred. No. 0.00078;  
Matches 206; Conservative 0; Mismatches 239; Indels 0; Gaps 0;  
QY 171 TAATGCAATGATCTGATTCGAATAAAGATGGAGCTTCTTTATTTAAATTCCTCAAGTTAA 230  
Db 65 TAC 124  
QY 231 AGAAGCTGAGTATAAGCTTCAACTGATTTTAAATCTCTTTTTCAGCAACTACTACTAATGG 290  
Db 125 TAC 184  
QY 291 AGGGAGACATATGTAACCTAAAAAGATAGTCCGTCAGCAATGAGATTGCCACATGGCC 350  
Db 185 TAGTGTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 244  
QY 351 TAAATCTATATCAGCTAATATATACACCACTTTCCTACTGTCTAGTCAAAATGATGAG 410  
Db 245 TAC 304  
QY 411 TACTGAGGTTAATATGTTTCCCAATATGGATATTATATGTTTCTAGCACCTGTTAATAA 470  
Db 305 TAC 364  
QY 471 TGGAGCTGTAATATGTTTACATCTGTAACCTCAATGCTACTATTTCATGAAAAGAAATAC 530  
Db 365 TACTGC 424  
QY 531 TGATGCGACATGGGAGATGGTGGGAAAACTGTAGATCAAAAAACGTACTCGGTGCG 590  
Db 425 TCGTGTCTACTACTACTACTACTGCTGCTGCTACTACTACTACTGCTACTACTACTACTGC 484  
QY 591 TGATCAGTCAAAATATATCTATTACT 615  
Db 485 TGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 509

#### RESULT 16

BX414650/c  
LOCUS  
DEFINITION  
994 bp mRNA linear EST 15-MAY-2003  
BX414650 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YN02  
3-PRIME, mRNA sequence.  
BX414650  
ACCESSION  
BX414650.1 GI:30763455  
VERSION  
EST.  
KEYWORDS

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6015.f Contact : Feng Liang Email : fliang@lifetech.com URL : <a href="http://fulllength.invitrogen.com/">http://fulllength.invitrogen.com/</a> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSOCAP001DG01NP1. Location/Qualifiers
FEATURES	1..994 <code>/organism="Homo sapiens"</code> <code>/mol_type="mRNA"</code> <code>/db_xref="taxon:9606"</code> <code>/clone="CSOCAP001YN02"</code> <code>/tissue_type="THYMUS"</code> <code>/clone_lib="Homo sapiens THYMUS"</code> <code>/note="vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."</code>
ORIGIN	Query Match Best Local Similarity     4.1%;   Score 61.8; DB 13; Length 994; Matches 138; Conservative 197; Mismatches 255; Indels 8; Gaps 1;
Qy	330 AAATGAGATTGGCAGACATGGGCCTAAACTATATCATGCTAATACTACTACACCAGTTCCACTGT 389
Db	968 AARDRWAGAARARRADWADWDAAAWRWRDWTAWWAARKKWKADTTTKTKKDADA 909
Qy	390 TACTGACTCAATAATGATGGTACTGAGGTATTATTAATGTTTCCCAATATCGATATATTATTA 449
Db	908 DWGDDDWMDATAWWAADWWAAARWAAAATWWWTDWTAWTRKDKATWKADDADDDDAKT 849
Qy	450 TGTTTTCTAGCACCTGTAATGAATGAGCTGTAATATGTTGTTACATCTCTGTAATCTCAAATGC 509
Db	848 TTTTITTTWRAWTWDAGTWAKWADATWADADAAARAARWAAARWAAAAWWAAADAD 789
Qy	510 TACTATTTCATGAAAAGAAATCTGATCGCATGGGAGATGGTGTTGGAAAAAACTGTGAGA 569
Db	788 WWWWWWDAADAAMWAAWDTKDKRAAADWARAAAAAAWWDWGRGGRADAAGDKDARAA 729
Qy	570 TCAAAAAACGTACTCGTGGTGATACAGTCAATATATCTATTACTTATTAAGAATGCAGT 629
Db	728 WWDGAGGRDKDRDWAANRDWDADAAARBARAAAAATAWWATWKKDKDXMAAAA----- 676
Qy	630 CAATTATCTAGTGCACAGAAAAGTGTATCAATATGTTATAAGGATACTATGCCATCTGC 689
Db	675 -AWDTAGKRGKDARDWNAWNTTTTTTTTAAWAADKAWKWKATATTTTTTTTTTTTTTT 617
Qy	690 TTCTGTAGTTGATTGGAACGAAGGCTCTATGAGTAGTAATATTACTGATGGATCAGGGAA 749
Db	616 TTTTKTKTTAKMTDWDADWTTTTTTTTTTTTTTTTTTTAAATTTTKTKTKTKTKKTTTTT 557
Qy	750 TATTACAACCTAACTCAGGTTTCGGAAAAAGCAACTCGGGAAGTATAAACCCTGTGATGAGGA 809
Db	556 TRDVTDAARARRRAA 497
Qy	810 AAAATAAATTTTACAGATTACTATTTCGTGGGAGCTACCAATACTCCAACCGGAAATAC 869
Db	496 ARRAANNRRDRNRNNIKTKWAWTRNKRGRTKTGTTTTTTTTTTDTKDQTGGKGKGTGR 437
Qy	870 TCAAAATGGAGCTAATGATGACTTTTTTTTTATAAGGGGAATAAATACATACATGACAGTCACTTA 929

```

LOCUS      PR0025683
DEFINITION F.rubripes GSS sequence, clone 450 bp DNA linear GSS 10-DEC-1997
ACCESSION AL018519
VERSION    AL018519.1 GI:2684887
KEYWORDS   GSS; genome survey sequence.
SOURCE     Takifugu rubripes (Fugu rubripes)
ORGANISM   Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Actinopterygii; Perciformes; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE  1 (bases 1 to 450)
AUTHORS    Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrana,Y.,
            Williams,G. and Brenner,S.
TITLE      Direct Submission
JOURNAL    Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
COMMENT     Centre Hinxton, Cambridge, CB10 1BS. Email: biohelp@hmp.mrc.ac.uk
            Vector: pBluescript II KS
            V type: phagemid
            PRIMER: KS
DESCR:     One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES   source
            1..450
            /organism="Takifugu rubripes"
            /mol_type="genomic DNA"
            /db_xref="taxon:31033"
            /clone_lib="154E17aC12"
            /clone_lib="cosmid 154E17"
ORIGIN
Query Match      4.0%; Score 61; DB 29; Length 450;
Best Local Similarity 49.7%; Pred. No. 0.0016; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 153;
QY 236 CTGAGTATAAGCTTCAACTGATTAAATCTCTTTTACGACAACTACTAATGGAGGA 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 138 CTGCTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 296 GACATATGTAACATAAAGATCTGCGTCAGCAATGAGATGGCATGGGCTAAAT 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 198 CTGCTGCTGCTACTGCTACTATTACTACTGCTACTACTGCTGCTACTACTGCTACTA 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 CTATATCAGCTAATACACACAGTTTCCACTGTTACTGAGTCAAAATATGATGGTACTG 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 258 CTACTGCTACTACTGCTACTGCTGCTACTGCTGCTACTGCTGCTACTGCTACTGCTA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 416 AGGTATTAATGTTTCCAAATATGGATATTATTATGTTTCTAGACATGTTAATAATGGAG 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 318 CTATTACTGCTACTACTGCTGCTGCTACTACTGCTACTACTACTGCTACTGCTGCTG 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 476 CTGTAATATGTTTACTGTAATCTCAATGCTACTATTCTCATGAAGAATGATGATG 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 378 CTGCTACTGCTACTACTACTGCTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTG 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 CGAC 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 438 CGAC 441

RESULT 19
BH211305/c
LOCUS      BH211305
DEFINITION Sml-50G22.TF Sml Schistosoma mansoni genomic clone Sml-50G22,
            genomic survey sequence.
ACCESSION BH211305
VERSION    BH211305.1 GI:16391571
KEYWORDS   GSS.
SOURCE     Schistosoma mansoni
ORGANISM   Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

```

---

```

REFERENCE  1 (bases 1 to 479)
AUTHORS    Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and
            El-Sayed,N.M.
TITLE      Use of end sequences from Schistosoma mansoni (Puerto Rico strain)
            Sml BAC library for gene discovery and map construction
JOURNAL    Unpublished (2001)
COMMENT     Other GSSs: Sml-50G22.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Schistosoma mansoni (Puerto Rico) Sml
            BAC library. For clone availability, please contact Dr. Najib
            El-Sayed at tigr (nelsayed@tigr.org) or Dr. Phillip LoVerde at
            State University of New York, Buffalo, New York, USA
            (loverde@buffalo.edu)
            Seq primer: M13 For
            Class: BAC ends.
FEATURES   Location/Qualifiers
            1..479
            /organism="Schistosoma mansoni"
            /mol_type="genomic DNA"
            /strain="Puerto Rico"
            /db_xref="taxon:6183"
            /clone="Sml-50G22"
            /clone_lib="Sml"
            /note="Vector: pBelosBAC11; Site 1: Hin dIII; Constructed
            in the laboratory of Dr. Denis Le Paslier at the Fondation
            Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma
            mansoni agarose embedded DNA was partially digested with
            Hin dIII. High molecular weight fragments were ligated in
            pBelosBAC11 digested with Hin dIII. The average insert size
            is 100 kb. Total clone coverage: approx. 7.95 x the
            haploid genome. Further information can be found in Le
            Paslier et al. (2000) Construction and characterization of
            a Schistosoma mansoni bacterial artificial chromosome
            library. Genomics 65: 87-94."
ORIGIN
Query Match      4.0%; Score 60.8; DB 28; Length 479;
Best Local Similarity 47.4%; Pred. No. 0.002; Indels 0; Gaps 0;
Matches 182; Conservative 0; Mismatches 202;
QY 362 CAGCTAATACTACACAGTTTCCACTGTTACTGAGTCAAAATATGATGCTAGGTTA 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 452 CTGCTGATGCTGCTGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGATG 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 422 TTAAATGTTTCCCAATATGATATTATTATGTTTCTAGCACCTGTTAATATGAGCTGTA 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 392 ATGATGATGCTGCCGCTGCTTATGATGATGATGATGATGATGATGATGATGATG 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 482 TTATGTTTACATCTGTAATCTCAATGCTACTATTTCATGAAAGAAATATCTGATGACAT 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 332 ATGATCTCTGCTGCTGCTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 542 GGGGAGATGGTGGTGGAAAAAAGTGTAGATCAAAAAACGTTACTCGTTGGTGTACAGTCA 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 272 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 602 AATATACTATTACTTATAAGATGAGTCAATATATCATGGTACAGAAAAGTGTATCAAT 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 212 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 662 ATGTTATAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 152 ATAAGAAAAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 722 AGTAACTATTACTGATGGATCAG 745
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 92 ACAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

QY 320 CTGGCTGAGCAATGAGATTGGACATGGCTAAATCTATATCAGTCACTAATCTACACGAG 379  
DB 249 CTA 308  
QY 380 TTTCCTACTGTTACTGAGTCAATTAATGAGTGGTACTGAGGTTAATGTTTCCCAATATG 439  
DB 309 CTA 368  
QY 440 GATATTATTATGTTCTAGACATGTTTAATTAATGAGGCTGTAATTAATGTTTCAATCTGTA 499  
DB 369 CTA 428  
QY 500 CTCCAATGCTACTATTCTAGAAAGCAATGATGATGGAC 539  
DB 429 CTA 468

RESULT 22  
CNS04AEE/c  
LOCUS  
DEFINITION  
Tetraodon nigroviridis genome survey sequence T7 end of clone  
095G24 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
ACCESSION  
AL281759.1 GI:8020086  
VERSION  
GSS; genome survey sequence.  
KEYWORDS  
Tetraodon nigroviridis  
ORGANISM  
Tetraodon nigroviridis

REFERENCE  
AUTHORS  
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
TITLE  
Estimate of human gene number provided by genome-wide analysis  
using Tetraodon nigroviridis DNA sequence  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Nat. Genet. 25 (2), 235-238 (2000)  
20296633  
10835645

TITLE  
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,  
Saurin,W., Bernot,A. and Weissenbach,J.  
TITLE  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Genome Res. 10 (7), 939-949 (2000)  
20359837  
10899143  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/tetraodon.  
FEATURES  
source  
Location/Qualifiers  
1. 806  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="095G24"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COBG095BD12LP1-end : T7"

ORIGIN  
Query Match 4.0%; Score 60; DB 29; Length 806;  
Best Local Similarity 41.6%; Pred. No. 0.0031;  
Matches 218; Conservative 22; Mismatches 284; Indels 0; Gaps 0;

QY 156 TGATGCAGAAATAGATAATGCAATGTATCTGATTCGAATAAAGATGGAGCTTCTTTATTT 215  
DB 753 TATTAATCAAT 694  
QY 216 AATTCCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTAC 275  
DB 693 TACTACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 634  
QY 276 GACAACTACTAATGGAGGGAACATATGTAACCTAAAAAGATACCTCGTCAGCAATGA 335  
DB 633 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 574  
QY 336 GATTGCGACATGGGCTAAATCTATATCAGCTAATATCTACACCACTTTCCACTGTTACTGA 395  
DB 573 TAATAATGGTAAATGTAAMSSTATTAAATVATACAAHMACWMMCMSCMCMCMCMCA 514  
QY 396 GTCAAAATAATGATGGTACTGAGGTTTATTAATGTTTCCCAATATGATATATATGTTTC 455  
DB 513 YMATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGGTAT 454  
QY 456 TAGCACTGTTAATAATGAGCTGTAATTTATGTTTACATCTGTAACCTCAAAATGCTACTAT 515  
DB 453 TAATGATAGTAAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 394  
QY 516 TCATGAAAAGAATACTGATGCGACATGGGAGATGGTGGTGGAAAAAACTGTAGATCAAAA 575  
DB 393 TAAATGGTAAATAATGATGTTAATGATGATGATGATGATGATGATGATGATGATGATG 334  
QY 576 AACGTACTCGGTTGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 635  
DB 333 CAACCCACCMATWATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 274  
QY 636 TCATGGTACAGAAAAAGTGTATCAATATGTTTATAAGGATACTA 679  
DB 273 TAACGGTATTATGATAGTAAATGATAATCATAGTAAATAATAATA 230

RESULT 23  
CNS01VTG  
LOCUS  
DEFINITION  
Tetraodon nigroviridis genome survey sequence T7 end of clone  
199C24 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
ACCESSION  
AL169549  
VERSION  
GI:7807606  
KEYWORDS  
GSS; genome survey sequence.  
SOURCE  
Tetraodon nigroviridis  
ORGANISM  
Tetraodon nigroviridis

REFERENCE  
AUTHORS  
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
TITLE  
Estimate of human gene number provided by genome-wide analysis  
using Tetraodon nigroviridis DNA sequence  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Nat. Genet. 25 (2), 235-238 (2000)  
20296633  
10835645

TITLE  
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,  
Saurin,W., Bernot,A. and Weissenbach,J.  
TITLE  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Genome Res. 10 (7), 939-949 (2000)  
20359837  
10899143  
Genoscope.

TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : <a href="mailto:secref@genoscope.cns.fr">secref@genoscope.cns.fr</a> ) - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/tetraodon">http://www.genoscope.cns.fr/tetraodon</a> .
FEATURES	Location/Qualifiers
source	1..773 /organism="Tetraodon nigroviridis" /mol_type="genomic DNA" /db_xref="taxon:99883" /clone="199C24" /clone_lib="G" /note="Genoscope sequence ID : COAG199BB121P1-end : T7"

ORIGIN	Query Match	4.0%; Score 59.8; DB 29; Length 773;
	Best Local Similarity	41.7%; Pred. No. 0.0035;
	Matches 178; Conservative 26; Mismatches 223; Indels 0; Gaps 0	
Qy	200	ATGAGGCTTCTTATTTTAATTCCTCAAGGTAAAGAAGCTGAGTATAAAGCTTCAACTGATT 259
Db	136	ATCATGTAAATTTTATTTTACTCATGATACTGCTACTACTACWACTGCTACWACTAGTG 195
Qy	260	TTAATTCCTTTTTACGCAACTACTAATGAGGGAGAACATATGTAATCTAATAAAAAAGATA 319
Db	196	CWACWAGCTACTGCACTACTGCTACTACTAGTGTACTACTGCTACTGCTACTACTG 255
Qy	320	CTGGGTGAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAGCTAATACTACACAG 379
Db	256	CTACTACWAGTGCTACTACTGCTACTACTACWACTGCACTGCTGCTACTACTGCTACTA 315
Qy	380	TTTCCACTGTTTACTGAGTCAAAATGAATGGTACTGAGGTTATTAAATGTTTCCCAATATG 439
Db	316	CTRCWACTACTACWACTGCTACTACTACTACTCTACTACTACTACTGCAWACCAGCATA 375
Qy	440	GATATTATTGTTTCTAGCACTGTTTAATATGGAGCTGTTAATTATGGTTACATCTGTAA 499
Db	376	CATACTACTGCTACTACTACATGCTACATCTACTGCAACTACWCAACTGCAWCTCTACAG 435
Qy	500	CTCCAAATGCTACTATTTCATGAAAGAATACTGATCGCAGATGGGGAGATGGTGGTGGNA 559
Db	436	CWACTACTACCACWACAACAACCTACWRCWCACTGCTACTGCTCAACTACTACTRCWA 495
Qy	560	AAACTGTAGATCAAAAAAGTACTCGGTTGGTGTGATACAGTCAAAATATATCTATTCTTATA 619
Db	496	CTGCTACTACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTGCTA 555
Qy	620	AGAATGC 626
Db	556	CTACTGC 562

RESULT	24
BJ3911130	
LOCUS	
DEFINITION	linear mRNA
ACCSSION	617 bp
VERSION	Dictyostelium discoidium cDNA library, SF Dictyostelium
KEYWORDS	discoidium cDNA clone dds15o06 5', mRNA sequence.
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE	Full length cDNA of Dictyostelium discoidium at the slug stage
JOURNAL	(bases 1 to 617)
COMMENT	Eukaryotes; Mycetozoa; Dictyosteliida; Dictyostelium. Unpublished (2002) Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

ORIGIN	Query Match	3.9%;	Score 59.6;	DB 12;	Length 617;
	Best Local Similarity	47.4%;	Pred. No. 0.0038;		
	Matches 179;	Conservative 0;	Mismatches 199;	Indels 0;	Gaps 0;
Qy	317	ATACTGCGTCAGCAAAATCAGATTCCGCACATGGGCTAAATCTATATCAGCTAATACTACAC	376		
Db	12	ATACTCAAACTACTACTACAAACAACAATAATAATAATAATAATAATAATAATAATAATAA	71		
Qy	377	CAGTTTCCACTGTTTACTCAGTCAAAATGAATGGTGACTTGAGGGTATTAAATGTTTCCCAAT	436		
Db	72	ATATAATGATTGATACAAATGACTATAAATGATGCTAGTGTCTTCAACCTCTTCTACAG	131		
Qy	437	ATGGATATTATTATGTTTCTAGCACTGTTTAATTAATGGAGCTGTTAATTTATGGTTACATCTG	496		
Db	132	TTGCTGAAGATTGTTCAAGTACTACTATTGATGTGTACAGCCACTATTTCAGATACAACAA	191		
Qy	497	TAACTCCAAATGCTPACTATTCTATTCAGAAAGAATACTGATCGCACATGGGAGATGGTGGT	556		
Db	192	CTAATCATAAAACATATGTAGATCAAGCAATGCAAGATAGAGAATGAGAGATCAAAATA	251		
Qy	557	GAAAAAAGCTGTAGATCAAAAAAGCTPACTCGGTTGGTGTGATACAGTCAAAATATATCTATTCTT	616		
Db	252	ATATAGCAATGATGATGATGATACAGATGAAATTTAATAATATAGATAAAAAATAAAGTAATA	311		
Qy	617	ATAGAATCGAGTCAAATTATTCATGGTACAGAAAAAGTGTATCAATATCTTTATAAAGGATA	676		
Db	312	ATATAGATAAAGANGAACAGAGCAAGAAAGAAAGAGAGAGATATACAAATACAA	371		
Qy	677	CTATGCCATCTGTTCTCTG	694		
Db	372	ATATAACATCAGATGATG	389		

RESULT 25	BZ422321	427 bp	DNA	linear	GSS 10-DEC-2002
LOCUS	BZ422321.c				
DEFINITION	id52c10.b1 WGS-SbicolorF (DH5a methyl filtered) Sorghum bicolor genomic clone id52c10 5', genomic survey sequence.				
ACCESSION	BZ422321				
VERSION	BZ422321.1	GI:26370493			
KEYWORDS	GSS.				
SOURCE	Sorghum bicolor (sorghum)				
ORGANISM	Sorghum bicolor				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.				
AUTHORS	1 (bases 1 to 427)				
TITLE	Rabinowicz, P.D., O'Shaughnessy, A.L., Ballija, V., Dedhia, N., Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L., Zutavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.				
JOURNAL	Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)				
COMMENT	Unpublished (2002)				
	Contact: W. Richard McCombie				
	Lita Annenberg Hazen Genome Sequencing Center				
	Cold Spring Harbor Laboratory				
	PO Box 100, Cold Spring Harbor, NY 11724, USA				
	Tel: 516 367 8984				







Db	154	GTTTTGTGACGATGACGATGATG 129	
RESULT 29			
FR0006944			
LOCUS			
DEFINITION		619 bp DNA linear GSS 02-MAR-1997	
ACCESSION		F.rubripes GSS sequence, clone 133B16aC7, genomic survey sequence.	
VERSION	290754		
KEYWORDS	290754.1	GI:1867968	
SOURCE		GSS; genome survey sequence.	
ORGANISM		Takifugu rubripes (Fugu rubripes)	
REFERENCE			
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontoidea; Tetraodontidae; Takifugu.	
TITLE		1 (bases 1 to 619)	
JOURNAL		Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y., Williams,G. and Brenner,S.	
COMMENT		Direct Submission Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.ac.uk Vector: pBluescript II KS V type: phagemid PRIMER: KS	
FEATURES		DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.	
source			
1. .619		Location/Qualifiers	
/organism="Takifugu rubripes"			
/mol_type="genomic DNA"			
/db_xref="taxon:31033"			
/clone_lib="133B16aC7"			
/clone_lib="cosmid 133B16"			
ORIGIN			
Query Match	3.8%;	Score 58;	DB 29; Length 619;
Best Local Similarity	46.7%;	Pred. No. 0.0089;	
Matches 169;	Conservative	0; Mismatches 193;	Indels 0; Gaps 0;
Qy	156	TGATGCGAAGATAGATAATGCAATGATCTGATTCGAATAAGATGGAGCTTCATT	215
Db	14	TGCTACTANNACTGCTGCTACNGCTACTACNGCTNCTGCTACTAATGCTACTGCTGCTAC	73
Qy	216	AATTCCTCAAGGTAAGAGCTGAGTAAAGCTTCAAGCTGATTTTAAATTCCTTTTAC	275
Db	74	TGCNACTACTGCTACTGCTGCTACTACTGCTGCTGCTGCTACTACTACTGCTGCTG	133
Qy	276	GACAACTACTAATGGAGGAGAACATATGTAATTAACAAAAAGATATGCGTCAGCAATGA	335
Db	134	TGCTGCTGCTGCTGCTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	193
Qy	336	GATTGCGACATGGGCTAAATCTATATGAGCTAATATACACAGTTTCCAGCTGTTACTGA	395
Db	194	TACTACTACTGCTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTACTAC	253
Qy	396	GTCAATAATGATGCTACTGAGGTTTAAATGTTTCCCAATATGGAATTAATTAATGTTTC	455
Db	254	TACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	313
Qy	456	TAGCACTGTTTAAATGGAGCTGAATTAATGTTTACATCTGTAACTCCAAATGCTACTAT	515
Db	314	TACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	373
Qy	516	TC 517	
Db	374	TC 375	
RESULT 30			
CE682324			
LOCUS			
DEFINITION		641 bp DNA linear GSS 29-SEP-2003	
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
1. .641		Location/Qualifiers	
/organism="Canis familiaris"			
/mol_type="genomic DNA"			
/strain="Standard Poodle"			
/db_xref="taxon:9615"			
/clone_lib="Dog Library"			
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"			
ORIGIN			
Query Match	3.8%;	Score 57.6;	DB 29; Length 641;
Best Local Similarity	48.7%;	Pred. No. 0.011;	
Matches 187;	Conservative	0; Mismatches 194;	Indels 3; Gaps 1;
Qy	153	TTTTCATGCGAAGATAGATAATGCAATGATCTGATTCGAATAAGATGGAGCTTCCTTA	212
Db	261	TGTTACTGCGATAGTGAATGCTACAAATCTACTGCTGCTGTTATTATTAATGCTATGTA	320
Qy	213	TTTAATTCCTCAAGTAAAGAGCTGAGTATAAGCTTCAAGCTTGAATTTTAAATTCCTTTT	272
Db	321	TGCTACAGCTACTGCTAATGCTGCTGCAATAACACTAATACCACTGCTGCTGCTATCAT	380
Qy	273	TACGCAACTACTAATGAGGGAGAACATATGTAATAAAAAAGATATGCGTCAGCAAA	332
Db	381	TACTAGTGTAGTGTGATGCTACTACCATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT	440
Qy	333	TGAGATTGCGACATGGGCTAAATCTATATCAGCTAATACACCAAGTTTCCAGCTGTTAC	392
Db	441	TAAATGCTACTAATGCTACTGCTACCACTCATATGCTAGTGTGCAACTCCTGAAAT	500
Qy	393	TGAGTCAATAATAGTGTGCTGAGGTTTAAATGTTTCCCAATATGATATTAATGCT	452
Db	501	ATCTAATACTAGTGTGCTAATGCTGCTAATACTGCTGCTAATACTGCTGCTGCTGCTGCT	560
Qy	453	TTCTAGCACTGTTAATTAATGAGCTGTTAATGTTTACATCTGTAATCTCCAAATGCTAC	512
Db	561	TACTGCTGATTAATCAATGCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	617
Qy	513	TATTCATGAAAGAACTACTGATGC	536
Db	618	TGGAATACACTAATACTACTGTC	641
Search completed: July 27, 2004, 01:14:14			
Job time : 4191 secs			

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 22:21:53 ; Search time 124 Seconds  
(without alignments)  
6753.394 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaaagaatgattca.....gtcgtcgttactgtcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64.8	4.3	1980	4	US-09-134-000C-2719
2	64.8	4.3	1983	4	US-09-071-035-217
3	56.2	3.7	1575	4	US-09-543-681A-615
4	50.4	3.3	14066	4	US-09-601-198-56
5	49.8	3.3	1687	4	US-09-071-035-219
6	48.4	3.2	3666	2	US-08-682-517-13
7	48.4	3.2	3666	2	US-08-682-517-14
8	48.4	3.2	4197	2	US-08-682-517-7
9	48.4	3.2	4197	2	US-08-682-517-8
10	47.8	3.2	740	3	US-09-451-117-1
11	47.8	3.2	740	4	US-09-888-655-1
12	47.4	3.1	1037	4	US-09-181-585-3
13	47.4	3.1	1159	4	US-09-181-585-1
14	47.4	3.1	1471	4	US-09-181-585-2
15	47.4	3.1	3945	4	US-09-200-6508-6
16	47.4	3.1	4358	4	US-08-956-171E-454
17	47.2	3.1	7218	1	US-08-232-463-14
18	46.8	3.1	1948	1	US-07-849-438-1
19	46.8	3.1	4590	4	US-09-134-001C-1108
20	45.6	3.0	4376	1	US-08-119-125A-1
21	45.6	3.0	6744	1	US-08-119-125A-2
22	45	3.0	832	4	US-09-621-976-2813
23	44.6	3.0	11679	4	US-09-328-352-1377
24	44.2	2.9	606	4	US-09-601-198-166
25	44.2	2.9	30549	4	US-09-134-001C-322
26	44	2.9	1716	4	US-09-134-001C-1028
27	43.4	2.9	2763	4	US-09-463-402-5

ALIGNMENTS

RESULT 1

US-09-134-000C-2719  
; Sequence 2719, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:

Sequence 3, Appli  
Sequence 13, Appl  
Sequence 10, Appl  
Sequence 84, Appl  
Sequence 84, Appl  
Sequence 55, Appl  
Sequence 86, Appl  
Sequence 86, Appl  
Sequence 1789, Ap  
Sequence 1, Appli  
Sequence 31, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 7, Appli  
Sequence 7, Appli  
Sequence 12, Appl  
Sequence 7, Appli  
Sequence 5, Appli  
Sequence 534, App  
Sequence 80, Appl  
Sequence 2, Appli  
Sequence 21, Appl  
Sequence 17, Appl  
Sequence 19, Appl  
Sequence 15, Appl  
Sequence 707, App  
Sequence 9, Appli  
Sequence 7110, Ap  
Sequence 21, Appl  
Sequence 1, Appli  
Sequence 321, Ap  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 954, App  
Sequence 24, Appl  
Sequence 24, Appl  
Sequence 24, Appl  
Patent No. 5352450  
Sequence 1, Appli  
Sequence 63, Appl  
Sequence 63, Appl  
Sequence 22, Appl  
Sequence 22, Appl  
Sequence 22, Appl  
Sequence 108, App  
Sequence 108, App  
Sequence 1, Appli  
Sequence 2509, Ap  
Sequence 1473, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1848, Ap  
Sequence 1887, Ap  
Sequence 1887, Ap  
Sequence 1, Appli  
Sequence 1010, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 487, App  
Sequence 3029, Ap  
Sequence 1575, Ap  
Sequence 841, App  
Sequence 1714, Ap  
Sequence 7, Appli



Db 464 TTAATGATCAAAATTTTCTGAGTGTATATATGAATTAGATATAAATAACAGTCTTCTGGT 523  
QY 227 GTAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTATCGACAACACTACTA 286  
Db 524 AAATAGACTTAAAGCTTAAATAAAGTAACTTTAAATCAAGTTAACTTTAGTCAATAAGAC 583  
QY 287 ATGGAGGAGACATATGTAATCAATAAAGATACCTGCGTCAGCAATGAGATTGGACAT 346  
Db 584 AGCTCGTAGCAAGCAAGGATAGAAAATGATTAATTAATTAATTAATTAATTAATTAATTA 643  
QY 347 GGGCTAAATCTATATCAGCTAATCTACACCAAGTTTCCACTGTTACTGAGTCAAAATAATG 406  
Db 644 TTAACCTAGAAAATGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 703  
QY 407 ATGGTACTGAGGTATTAATGTTTCCCAATATGGAATATTAATTAATTAATTAATTAATTA 466  
Db 704 ATATTAACTTAGAAAATGATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 763  
QY 467 ATAATGGAGCTGTAATTAATGTTTCCCAATATGGAATATTAATTAATTAATTAATTAATTA 526  
Db 764 ATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 823  
QY 527 ATACTGATCGGACATGGGAGATGGTGGGAAAACCTGATCAAAAACGCTACTCG 586  
Db 824 ATGATAATGTTAACTTAGAAAATGATATTAATTAATTAATTAATTAATTAATTAATTA 883  
QY 587 TTGGTGATCAGCTCAAAATATCTATTACTTTAAGAATGCACTCAATTAATCAATGATGTA 646  
Db 884 AAAATGATAATGTTAACTTAGAAAATGATATTAATTAATTAATTAATTAATTAATTAAT 943  
QY 647 AAAAGTGTCAATATGTTTAAAGGATACATATGCGCATCTGCTCTGTAGTTGATTTGA 706  
Db 944 TAAATATAAAAAGGCACTAATAAAGATATAAATAGAAATCTTATATTTTCAATGGTA 1003  
QY 707 ACCAA 711  
Db 1004 AAGAA 1008

## RESULT 4

US-09-601-198-56/c

; Sequence 56, Application US/09601198

; Patent No. 6531583

; GENERAL INFORMATION:

; APPLICANT: Cassell, Gail H.

; APPLICANT: Chen, Ellison Y.

; APPLICANT: Glass, Jennifer S.

; APPLICANT: Glass, John I.

; APPLICANT: Heiner, Cheryl R.

; APPLICANT: Lefkowitz, Elliot

; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA

; FILE REFERENCE: UAB-13452/22

; CURRENT APPLICATION NUMBER: US/09/601,198

; PRIOR FILING DATE: 2000-12-08

; PRIOR FILING DATE: 1998-01-30

; NUMBER OF SEQ ID NOS: 181

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 56

; LENGTH: 14066

; TYPE: DNA

; ORGANISM: Ureaplasma urealyticum

US-09-601-198-56

Query Match 3.3%; Score 50.4; DB 4; Length 14066;

Best Local Similarity 42.8%; Pred. No. 0.007;

Matches 313; Conservative 0; Mismatches 416; Indels 3; Gaps 1;

QY 435 ATATGATATTTATGTTTCTAGCACTGTTAATGAGCTGTAATTTATGTTTACATC 494

Db 5284 AAATGTTAGTTCTAATGATCTTATAAATACTACACACACAGTTAATTTGATTTGA 5225

QY 495 TGTAACTCCAATGCTACTATTTCATGAAAAGAAATCTGATGGGACATCGGGAGATGGTGG 554  
Db 5224 TGGAAATCAACGAACCTTGAATTAATAAAAAAATTCATTTGGTATATCTTCAAAATGACGG 5165  
QY 555 TGGAAAACTGATGATCAAAAAACGCTACTCGTGGTGGTATACAGTCAAAATATATCTATTAC 614  
Db 5164 TGAAGAAATTTTAAAGTATCAAAAAACATTCGCTTTGGCTAATTAACCACTATAGTTTGA 5105  
QY 615 TTATAAGAAATCAGCTCAATTTATCATGTTGTAAGAAAAAGTGTATCAATATGTTTAAAGGA 674  
Db 5104 ATTGCTAAATTTAAAAACATAATCGTAAATATACATAAAAAAGAAATTAATAATGA 5045  
QY 675 TACTATGCCATCTGCTGCTGCTAGTTGATTTGAACGAAGGCTTTATGAAGTAACTATTAC 734  
Db 5044 TGATAATAAAAAACATCAATTAATTTTCCATTTAAAAAATGGAATAGCTGATTTATTTGT 4985  
QY 735 TGATGATCAGGAATATTTCAAACTCAAACTCAAGTTTCGGAAGAAAGCAACTGGGAAGTA 794  
Db 4984 TAATAAACAACATCGATTTAGTTTAAGTTCAATTAATGAGCAATTTGCAAGGCGAAAAA 4925  
QY 795 TAACCTGTTAGAGGAAAAATAAATTTTCAAGATTAATTTCCGTGGGCGAGCTACCAATAC 854  
Db 4924 TAATTTACAATCAACACAAATTAGATTTTATTAATGATCTGTATAATGTTTAAAGTAA 4865  
QY 855 TCCAAACCGAAATCTCAAAATGGAGCTAATGATGACTTTTTTATTAAGGGAATTAATAC 914  
Db 4864 TGAAGAAGAACAATCAATTAATTTATGTTGAGAAAT---TTATCTGTAAAAAGCAAAAGTTAA 4808  
QY 915 AATCAGCTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCCAGGTTCAAGCTGATTT 974  
Db 4807 AATTTAGGGAGTCAAAAAATTTTAGAAAAACAATTTCAATTAATCTAGCAATTAACCAAGA 4748  
QY 975 ACCAGAAATACAAACATTTGGGACCATCAACCCCAATACTAGCAATGATGACCCAGTCA 1034  
Db 4747 TACTATTATTAACAGTATTACTTTTATTAATAAACCAATAAAAGCTGTACGAATTTGG 4688  
QY 1035 AAAAGTAAAGTGGGAGTGGTCAAAATTTACTATAAAAAAATTTGATGTTTCCACAAAAAGC 1094  
Db 4687 AATAGATAATAGCAATAAATAATATATAACAATAACAATAAATTAATTAATTAATTAATTA 4628  
QY 1095 TTCATTACAAGTGTCTATTTTGTGTTTAAAGAAATGCTACGGTCAATTTCTTAAACCTTTAA 1154  
Db 4627 TAATAATTTTAGTGTCAATGTTCTGTTCAAAAGTACTAATAAAGCTATTAACGTTATTTGGA 4568  
QY 1155 CGATACAAATAA 1166  
Db 4567 ATATAATAATAA 4556

## RESULT 5

US-09-071-035-219

; Sequence 219, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; FILING DATE: US/09/071,035

; CLASSIFICATION:



QY 542 GGGGAGATGGTGGTGGAAAAAAGTGTAGATCAAAAAAGCTACTCGGTTGGTGATACAGTCA 601  
DB 3008 GAGCAATTAATCTTCTGATACATTTACACAAGGTGTATTACCATCAGCAGCTACAGCAGCTG 3067  
QY 602 AATATACATTAATCTTATTAAGATGAGTCAATTTATCATGGTACAGAAAAAGGTATCAAT 661  
DB 3068 AATATACATTAATCTTATTAAGATGAGTCAATTTATCATGGTACAGAAAAAGGTATCAAT 3127  
QY 662 ATGTTTATAAGGATACATATGCCATCTGCTTCTGTAGTTGATTTGAACGAAGGGTCTTATG 721  
DB 3128 TAAATATTGATAATGCTGGTGTCTCAAGTAATTAACCTTACAGGTAATAAAGGTGCACAAG 3187  
QY 722 AAGTAATTAATTAAT 735  
DB 3188 GTGTAGCTGATGCT 3201

## RESULT 8

US-08-682-517-7  
; Sequence 7, Application US/08682517  
; Patent No. 5874267  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Expression of surface layer proteins  
; NUMBER OF SEQUENCES: 25  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/682,517  
; FILING DATE:  
; CLASSIFICATION:  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4197 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-682-517-7

Query Match 3.2%; Score 48.4; DB 2; Length 4197;  
Best Local Similarity 44.5%; Pred. No. 0.014;  
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;  
QY 302 ATGTAACCTAAAAAGATAGTGGTCAAGCAATGAGATTCGCGACATGGGCTAAATCTATAT 361  
DB 2952 ATGTAAGTGCAACAACCTGTTGATCTGCAACTGTTTCATTAAAGATAGTCAAAATAAT 3011  
QY 362 CAGCTAATACTACACCAGTTTCCACTGTTTACTGAGTCAAAATATGATGATGATGAGGTTA 421  
DB 3012 CATTATCTTCAATAGTTGAACTGGTGCTTAATACAGGTGATTTGCTCAACTGTTTC 3071  
QY 422 TTAATGTTTCCCAATATGATATTAATGTTTCTAGCACTGTTAATAATGAGCTGTAA 481  
DB 3072 AAGCTGGTACATTAATCTTCTTAACTGCTGGTACATTAACAGTACTTATGCGATGCTA 3131  
QY 482 TTATGGTTTACATCTGTAACCTCAAAATGCTACTTATTCATGAAAAAGAAATACTGATGCGACAT 541  
DB 3132 AAAATGCTGAGGTGTTGCTGAAAATATTACTGTAGCTAACATTAAAGAAACTACTG 3191  
QY 542 GGGGAGATGGTGGGAAAAAAGTGTAGATCAAAAAAGCTACTCGGTTGGTGATACAGTCA 601  
DB 3192 GAGCAATTAATCTTCTGATACATTTACACAAGGTGATTAACCATACAGCAGCTACAGCAGCTG 3251  
QY 602 AATATACATTAATCTTATTAAGATGAGTCAATTTATCATGGTACAGAAAAAGGTATCAAT 661  
DB 3252 AATATACATTAATCTTATTAAGATGAGTCAATTTATCATGGTACAGAAAAAGGTATCAAT 3311  
QY 662 ATGTTTATAAGGATACATATGCCATCTGCTTCTGTAGTTGATTTGAACGAAGGGTCTTATG 721

DB 3312 TAAATATTGATAATGCTGGTCTCAAGTAATTAACCTTAGCAGGTAAAAAAGGTGCACAAG 3371  
QY 722 AAGTAACCTATTACT 735  
DB 3372 GTGTAGCTGATGCT 3385  
RESULT 9  
US-08-682-517-8  
; Sequence 8, Application US/08682517  
; Patent No. 5874267  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Expression of surface layer proteins  
; NUMBER OF SEQUENCES: 25  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/682,517  
; FILING DATE:  
; CLASSIFICATION:  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4197 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus sphaericus  
; INDIVIDUAL ISOLATE: P-1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 95..3850  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 185..3850  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 95..184  
US-08-682-517-8

Query Match 3.2%; Score 48.4; DB 2; Length 4197;  
Best Local Similarity 44.5%; Pred. No. 0.014;  
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;  
QY 302 ATGTAACCTAAAAAGATAGTGGTCAAGCAATGAGATTCGCGACATGGGCTAAATCTATAT 361  
DB 2952 ATGTAAGTGCAACAACCTGTTGATCTGCAACTGTTTCATTAAAGATAGTCAAAATAAT 3011  
QY 362 CAGCTAATACTACACCAGTTTCCACTGTTTACTGAGTCAAAATATGATGATGATGAGGTTA 421  
DB 3012 CATTATCTTCAATAGTTGAACTGGTGCTTAATACAGGTGATTTGCTCAACTGTTTC 3071  
QY 422 TTAATGTTTCCCAATATGATATTAATGTTTCTAGCACTGTTAATAATGAGCTGTAA 481  
DB 3072 AAGCTGGTACATTAATCTTCTTAACTGCTGGTACATTAACAGTACTTATGCGATGCTA 3131  
QY 482 TTATGGTTTACATCTGTAACCTCAAAATGCTACTTATTCATGAAAAAGAAATACTGATGCGACAT 541  
DB 3132 AAAATGCTGAGGTGTTGCTGAAAATATTACTGTAGCTAACATTAAAGAAACTACTG 3191  
QY 542 GGGGAGATGGTGGGAAAAAAGTGTAGATCAAAAAAGCTACTCGGTTGGTGATACAGTCA 601  
DB 3192 GAGCAATTAATCTTCTGATACATTTACACAAGGTGATTAACCATACAGCAGCTACAGCAGCTG 3251  
QY 602 AATATACATTAATCTTATTAAGATGAGTCAATTTATCATGGTACAGAAAAAGGTATCAAT 661  
DB 3252 AATATACATTAATCTTATTAAGATGAGTCAATTTATCATGGTACAGAAAAAGGTATCAAT 3311



QY 662 ATGTTTAAAGATACATATGCCATCTGCTTCTGTAGTTGATTTGAACGAGGCTTTATG 721  
DB 3312 TAAATATTGATATGCTGGTGTCTCAAGTAATTAAGTACAGGTAAAAAAGGTGCACAAG 3371  
QY 722 AAGTAACTATTACT 735  
DB 3372 GTGTAGCTGATGCT 3385

RESULT 10  
US-09-451-117-1  
; Sequence 1, Application US/09451117  
; Patent No. 6277973  
; GENERAL INFORMATION:  
; APPLICANT: Jenkins, Mark C.  
; APPLICANT: Fayer, Ronald  
; APPLICANT: Trout, James  
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41  
; Patent No. 6277973  
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein  
; FILE REFERENCE: 0046.99  
; CURRENT APPLICATION NUMBER: US/09/451,117  
; CURRENT FILING DATE: 1999-11-30  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 740  
; TYPE: DNA  
; ORGANISM: Cryptosporidium parvum  
US-09-451-117-1

Query Match 3.2%; Score 47.8; DB 3; Length 740;  
Best Local Similarity 46.7%; Pred. No. 0.011;  
Matches 151; Conservative 0; Mismatches 172; Indels 0; Gaps 0;  
QY 217 ATTCCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAACGATTTTAAATTTCTTTTACG 276  
DB 370 ACTACTCAATCAATGGAAGTTACTGATCAATATTTGGTAATATGAGTGTATTATTACA 429  
QY 277 ACACTACTAATGAGGAGAACATATGTAACATAAAGATACGCTCAGCAATGAG 336  
DB 430 TCTAGTGGTATCTATAGCTGTTACTAATAATCTTAATGTAATAATAAGTAATAGT 489  
QY 337 ATTGGGACATGGGCTTAATCTATATCAGTAATACACCAAGTTTCCACGTGTACTGAG 396  
DB 490 AATATTGATCAGGAATTTTATACCAAGTTGGTACTTGTCTTCTACTAGTATGGTAAT 549  
QY 397 TCAAATAATGATGGTACTGAGGTTTAAATGTTTCCCAATATGGATATTTATATGTTTCT 456  
DB 550 AGTAATGGTGTGCTTTTACTGCTATTTCATCCTAATAATAACATAGCAATAATTAAT 609  
QY 457 AGCACTGTTAATATGAGCTGTAATTTATGGTTACATCTGTAACCTCCAAATGCTACTATT 516  
DB 610 AATAATAATAATAATAGTAATACCACTCTTACTACTGTTGCTACTAATGCTAATATT 669  
QY 517 CATGAAAAGAACTACTGATCGAC 539  
DB 670 ACTACTAATACTAATACTACTAC 692

RESULT 11  
US-09-888-655-1  
; Sequence 1, Application US/09888655  
; Patent No. 6521229  
; GENERAL INFORMATION:  
; APPLICANT: Jenkins, Mark C.  
; APPLICANT: Fayer, Ronald  
; APPLICANT: Trout, James  
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41  
; Patent No. 6521229  
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein  
; FILE REFERENCE: 0046.99

; CURRENT APPLICATION NUMBER: US/09/888,655  
; CURRENT FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 740  
; TYPE: DNA  
; ORGANISM: Cryptosporidium parvum  
US-09-888-655-1

Query Match 3.2%; Score 47.8; DB 4; Length 740;  
Best Local Similarity 46.7%; Pred. No. 0.011;  
Matches 151; Conservative 0; Mismatches 172; Indels 0; Gaps 0;  
QY 217 ATTCCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAACGATTTTAAATTTCTTTTACG 276  
DB 370 ACTACTCAATCAATGGAAGTTACTGATCAATATTTGGTAATATGAGTGTATTATTACA 429  
QY 277 ACACTACTAATGAGGAGAACATATGTAACATAAAGATACGCTCAGCAATGAG 336  
DB 430 TCTAGTGGTATCTATAGCTGTTACTAATAATCTTAATGTAATAATAAGTAATAGT 489  
QY 337 ATTGGGACATGGGCTTAATCTATATCAGTAATACACCAAGTTTCCACGTGTACTGAG 396  
DB 490 AATATTGATCAGGAATTTTATACCAAGTTGGTACTTGTCTTCTACTAGTATGGTAAT 549  
QY 397 TCAAATAATGATGGTACTGAGGTTTAAATGTTTCCCAATATGGATATTTATATGTTTCT 456  
DB 550 AGTAATGGTGTGCTTTTACTGCTATTTCATCCTAATAATAACATAGCAATAATTAAT 609  
QY 457 AGCACTGTTAATATGAGCTGTAATTTATGGTTACATCTGTAACCTCCAAATGCTACTATT 516  
DB 610 AATAATAATAATAATAGTAATACCACTCTTACTACTGTTGCTACTAATGCTAATATT 669  
QY 517 CATGAAAAGAACTACTGATCGAC 539  
DB 670 ACTACTAATACTAATACTACTAC 692

RESULT 12  
US-09-181-585-3  
; Sequence 3, Application US/09181585  
; Patent No. 6524791  
; GENERAL INFORMATION:  
; APPLICANT: Ranum, Laura P.W.  
; APPLICANT: Koob, Michael  
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION  
; FILE REFERENCE: 11000900101  
; CURRENT APPLICATION NUMBER: US/09/181,585  
; CURRENT FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 1037  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: cDNA  
; OTHER INFORMATION: comprising exons E, C, and A  
US-09-181-585-3

Query Match 3.1%; Score 47.4; DB 4; Length 1037;  
Best Local Similarity 47.2%; Pred. No. 0.016;  
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;  
QY 218 TTCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAACGATTTTAAATTTCTTTTACGA 277  
DB 611 TCCTTCATGTTAGAAAACCTGGCTTTTACTACTACTACTACTACTACTACTACTACTA 670  
QY 278 CAACCTACTAATGAGGAGAACATATGTAACATAAAGATACGCTCAGCAATGAGA 337  
DB 671 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 730



```
QY 162 AGAATAGATATGCAATGATCTGATTCGATTAAGATGAGCTTCTTTATTTATTC 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 333 AGAAAAAGAAAGTGTCAATCTACCACTGGAATAAAGTTGAAGTTTCAACTGCCAAATC 392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 222 TCAAGGTAAAGAGCTGAGTATATAAGCTTCAACTGATTTTAAATTTCTTTTACGACAC 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 393 AGATGAGCAGCTTACCAAAATCTACGAATGAAGATTTAAACACTAAACAACTATAAG 452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 282 TACTAATGAGGGAGAACATATGATATACTAAAAAGATCTGGCTCAGCAAAATGAGATTC 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 453 TAATCAAGAGCGCTTCAACCTGATTTGCAAGAGATAAATCAGTGGTAAATGTTCAACC 512
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 342 GACATGGGCTAAATCTATATCAGTAACTACTACACCACTTTCACCTGTTACTGAGTCAA 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 513 AACTAATGAGGAAACAAAGAGGTAGATGCCAAACTGAATCAACTACATTTAAATGTTAA 572
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 402 TAATGATGGTACTGAGGTTATTAAATGTTTCCCAATATGGATATTAATGTTTCTAGCAC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 573 AAGTGATGCTATCAAGAGTAATGATGAACCTCTTGTGATAACAATAGTAATTCAAATAA 632
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 462 TGTAAATAATGAGAGCTGATTAATGATGTTTACATCTGTAACCTCAA 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 633 TGAATAATGCGAGATATCATCTTTTGCCAAAAAGTACAGCACCTAA 677
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 16
US-08-956-171E-454/c
; Sequence 454, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
;           Gil H. Choi
;           Patrick S. Dillon
;           Craig A. Rosen
;           Steven C. Barash
;           Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 454:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; LENGTH: 4358 base pairs
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 454:
```

```
US-08-956-171E-454
Query Match 3.1%; Score 47.4; DB 4; Length 4358;
Best Local Similarity 46.1%; Pred. No. 0.026;
Matches 159; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
QY 162 AGAATAGATATGCAATGATCTGATTCGAATATAAGATGAGCTTCTTTATTTATTC 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 631 AGAAAAAGAAAGTGTCAATCTACCACTGGAATAAAGTTGAAGTTTCAACTGCCAAATC 572
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 222 TCAAGGTAAAGAGCTGAGTATATAAGCTTCAACTGATTTTAAATTTCTTTTACGACAC 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 571 AGATGAGCAGCTTCAACCAAAATCTACGAATGAAGATTTAAACACTAAACAACTATAAG 512
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 282 TACTAATGAGGGAGAACATATGATATACTAAAAAGATCTGGCTCAGCAAAATGAGATTC 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 511 TAATCAAGAGCGTTACACCTGATTTGCAAGAGATAAATCAGTGGTAAATGTTCAACC 452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 342 GACATGGGCTAAATCTATATCAGCTAAATACTACCACTGTTTCCACTGTTACTGAGTCAA 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 451 AACTAATGAGGAAACAAAGAGGTAGATGCCAAACTGAATCAACTACATTTAAATGTTAA 392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 402 TAATGATGGTACTGAGGTTATTAAATGTTTCCCAATATGGATATTAATGTTTCTAGCAC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 391 AAGTGATGCTATCAAGAGTAATGATGAACCTCTTGTGATAACAATAGTAATTCAAATAA 332
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 462 TGTAAATAATGAGAGCTGATTAATGATGTTTACATCTGTAACCTCAA 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 TGAATAATGCGAGATATCATCTTTTGCCAAAAAGTACAGCACCTAA 287
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 17
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F18
US-08-232-463-14

Query Match      3.1%; Score 47.2; DB 1; Length 7218;
Best Local Similarity 5.6%; Pred. No. 0.034;
Matches 22; Conservative 206; Mismatches 164; Indels 0; Gaps 0;

QY 885 TGATGACCTTTTATAGGGTAATAACATACATCACAGTCACCTTATACAGGAGTATTAA 944
Db 1482 TAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAAATTGGTACRRRRRRRR 1423
QY 945 GAGTGAGGCTAAACAGGCTGAGCTGATTTACGAGAAATACAAACATTCGCCACCATCA 1004
Db 1422 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1363
QY 1005 CCCAATACTAGCATGATGACCCAGGTCAAAAAGTAACAGTGAGGATGGTCAAAATTAC 1064
Db 1362 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1303
QY 1065 TATAAAAAAATTGATGGTTCACAAAAGCTTCATTACAAGTGTCTATATTGTTTAA 1124
Db 1302 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1243
QY 1125 GAATGCTACGGCTCAATTTCTAAACTTTAAACGATACAAATAACGTTGAATGGGCACAG 1184
Db 1242 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1183
QY 1185 AGCTAATGCAACAGATATACACAGGAGCAGATGGTATATACATTACAGGCTTGA 1244
Db 1182 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1123
QY 1245 AGAAGGTACATCTATCTAGTTGAGAAAAAG 1276
Db 1122 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1091
```

```
RESULT 18
US-07-849-438-1
; Sequence 1, Application US/07849438
; Patent No. 5459034
; GENERAL INFORMATION:
; APPLICANT: TABAOCHALI
; TITLE OF INVENTION: C. Difficile Specific Oligonucleotides
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 3i Research Exploitation Limited
; STREET: The Gate House, 2 Park Street
; CITY: Windsor
; STATE: Berkshire
; COUNTRY: U.K.
; ZIP: SL4 1LU
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/849,438
; FILING DATE: 19920519
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB90/02031
; FILING DATE: 28-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8929293.2
; FILING DATE: 29-DEC-1989
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1948 base pairs
```

```
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Clostridium difficile
US-07-849-438-1

Query Match      3.1%; Score 46.8; DB 1; Length 1948;
Best Local Similarity 45.4%; Pred. No. 0.027;
Matches 168; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 195 CTGATTCGAATAAGATGGAGCTTCTTATTTAACTCCTCAAGGTAAAGAAGCTGAGTATA 244
Db 794 CTGATCATAATAATATAGAGGTCAGGCTATAGTTTATCAAAGTAAATTTCTTAACTTTGA 853
QY 245 AAGCTTCAACTGATTTTAAATTTCTTTTTAGCAGCAACTACTAAATGAGGAGGAAACATATG 304
Db 854 ATGGTAAAAAATATTTATTTTGATATAATTTCAAAGCAGTTACTGATGGCAAACTATTG 913
QY 305 TAACCTAAAAAAGATAGTCCGTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAG 364
Db 914 ATAGTAAAAAATATTTACTTTTAATACTAACACACTGCTGAAGCAGCTACTGGATGGCAAACTA 973
QY 365 CTAATACTACACCAGTTTCCACTGTTACTAGTCAAAATAGATGAGTACTGAGGTTATTA 424
Db 974 TTGATGTTAAAAAATATTTACTTTTAATACTAACACACTGCTGAAGCAGCTACTGGATGGCAAA 1033
QY 425 ATGTTTCCCAATATGATTTATTTATTTCTAGCACCTGTTAAATAATGAGCTGTAAATTA 484
Db 1034 CTATTGATGGTAAAAAATTTACTTTTAATACTAACACCTGCTATAGCTTCACTGTTTATA 1093
QY 485 TGGTTACATCTGTAACCTCAAAATGCTACTATTTCATGAAAAAGAAATCTGATGCGACATGG 544
Db 1094 CAATTTATTAATGGTAAACATTTTATTTTAACTGATGTTATTCGAGTGTTTTAAAG 1153
QY 545 GAGATGGTGG 554
Db 1154 GACCTAATGG 1163
```

```
RESULT 19
US-09-134-001C-1108
; Sequence 1108, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lytun Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1108
; LENGTH: 4590
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1108

Query Match      3.1%; Score 46.8; DB 4; Length 4590;
Best Local Similarity 47.6%; Pred. No. 0.037;
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 289 GGAGGGAGAACATATGTAACTAAAAAAGATAGTGGTCAGCAAAATGAGATTCGACATGG 348
Db 2899 GAAGGTGGACAACTACCAAGGTTCAAAAGTATATCCATGATTCCTGCTGAGATAGAGTTTCG 2958
QY 349 GCTAAATCTATATCAGCTAATACACAGTTTCCACTGTCTACTGACTCAATATGAT 408
```

Db 2959 ACACGAGTATAGGATTAATTTCCACACACACACCATGATATTATTCAATTGAGGAC 3018  
QY 409 GGTACTGAGGTTAATGTTCCCAATATGGATATTATTGTTTCTAGCACTGTTAAT 468  
Db 3019 TTAGCACAGCTCATTCATGATTTTAAAAAATGCAATAGAGAGCTGATATTGCAAGTTAAG 3078  
QY 469 AATGGAGCTGTAATTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAGAAT 528  
Db 3079 CTTGTATCAAAAACCTGGCGTTGGAACATATAGCTTCAGGGGTAGCTAAAGCTTTCGCCGAT 3138  
QY 529 ACTGATGCGACATCGGGAGATGTTGGTGGAAAACTGTAGATCAAAAAAC 578  
Db 3139 AAAATTGTTATAAGTGGTTATGATGGAGGTACAGGTGCATCGCCTAAAAC 3188  
RESULT 20  
US-08-119-125A-1  
; Sequence 1, Application US/08119125A  
; Patent No. 5610011  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, Hilda Elizabeth  
; APPLICANT: VECHT, Uri  
; TITLE OF INVENTION: DNA Sequences which code for Virulence  
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, poly  
; TITLE OF INVENTION: Antibodies derived therefrom and the use thereof for the diagn  
; TITLE OF INVENTION: protection against infection by S. suis in mammals, including  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Centraal Diergeneeskundig Instituut  
; STREET: Edelhertweg 15  
; CITY: PH Lelystad  
; STATE:  
; COUNTRY: The Netherlands  
; ZIP: NL-8219  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS v. 6.0  
; SOFTWARE: WordPerfect v. 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/119,125A  
; FILING DATE: 20-SEP-1993  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/NL92/00054  
; FILING DATE: 19-MAR-1992  
; APPLICATION NUMBER: NL 9100510  
; FILING DATE: 21-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Handal, Anthony H.  
; REGISTRATION NUMBER: 26275  
; REFERENCE/DOCKET NUMBER: SMITHHE119125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (203) 838-8589  
; TELEFAX: (203) 838-8794  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4376 base pairs  
; TYPE: Nucleic acid with corresponding amino acids  
; STRANDEDNESS: single stranded  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus suis type II (pathogenic)  
; FEATURE:  
; OTHER INFORMATION: Extracellular protein factor (EF) gene  
; FEATURE:  
; NAME/KEY: promoter -35 region  
; LOCATION: bp 66 to 71  
; FEATURE:  
; NAME/KEY: promoter -10 region  
; LOCATION: bp 89 to 94  
; FEATURE:

; NAME/KEY: promoter -35 region  
; LOCATION: bp 153 to 158  
; FEATURE:  
; NAME/KEY: promoter -10 region  
; LOCATION: bp 176 to 181  
; FEATURE:  
; NAME/KEY: ribosome binding site  
; LOCATION: bp 350 to 356  
; FEATURE:  
; NAME/KEY: signal peptide  
; LOCATION: bp 361 to 498  
; FEATURE:  
; NAME/KEY: mature peptide  
; LOCATION: bp 499 to 2890  
; FEATURE:  
; NAME/KEY: dyad symmetry regions  
; LOCATION: from bp 4186 to 4198 and from bp 4203 to 4215  
; FEATURE:  
; NAME/KEY: dyad symmetry regions  
; LOCATION: from bp 4243 to 4257 and from bp 4263 to 4276  
; US-08-119-125A-1  
Query Match 3.0%; Score 45.6; DB 1; Length 4376;  
Best Local Similarity 51.5%; Pred. No. 0.072;  
Matches 105; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 365 CTAATACTACACAGTTCCTACTGTTACTGAGTCAAAATAATGATGCTACTGAGGTTATTA 424  
Db 3772 CTGTTATTAGCCAGGTACAGTTGTTACAGCATCTCTAAGACGAGGTTTGGAAATAGTT 3831  
QY 425 ATGTTTCCCAATATGATATTATTATGTTTCTAGCACTGTTTAATATGAGCTGTAAATTA 484  
Db 3832 CTGATTTCAGCAAGAGCTGAAGTTGTAGCAGTAGACGAAAAATAATTCTGCAGCAGGAGTGA 3891  
QY 485 TCGTTACATCTGTAACCTCAAAATGCTACTATTTCATGAAAGAAATACTGATCGACATGG 544  
Db 3892 AAGTTAAATCAGTTACTACAAATGCTAAATAATGTTGAGAAGAAAGCTAAGCAATTACCGA 3951  
QY 545 GAGATGGTGGTGGAAAACTGTAG 568  
Db 3952 ATACTGGTAGGAGNAGCAATTCAG 3975  
RESULT 21  
US-08-119-125A-2  
; Sequence 2, Application US/08119125A  
; Patent No. 5610011  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, Hilda Elizabeth  
; APPLICANT: VECHT, Uri  
; TITLE OF INVENTION: DNA Sequences which code for Virulence  
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, poly  
; TITLE OF INVENTION: Antibodies derived therefrom and the use thereof for the diagn  
; TITLE OF INVENTION: protection against infection by S. suis in mammals, including  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Centraal Diergeneeskundig Instituut  
; STREET: Edelhertweg 15  
; CITY: PH Lelystad  
; STATE:  
; COUNTRY: The Netherlands  
; ZIP: NL-8219  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS v. 6.0  
; SOFTWARE: WordPerfect v. 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/119,125A  
; FILING DATE: 20-SEP-1993  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/NL92/00054  
; FILING DATE: 19-MAR-1992  
; APPLICATION NUMBER: NL 9100510  
; FILING DATE: 21-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Handal, Anthony H.  
; REGISTRATION NUMBER: 26275  
; REFERENCE/DOCKET NUMBER: SMITHHE119125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (203) 838-8589  
; TELEFAX: (203) 838-8794  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4376 base pairs  
; TYPE: Nucleic acid with corresponding amino acids  
; STRANDEDNESS: single stranded  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus suis type II (pathogenic)  
; FEATURE:  
; OTHER INFORMATION: Extracellular protein factor (EF) gene  
; FEATURE:  
; NAME/KEY: promoter -35 region  
; LOCATION: bp 66 to 71  
; FEATURE:  
; NAME/KEY: promoter -10 region  
; LOCATION: bp 89 to 94  
; FEATURE:

Db 6320 ATACTGCTGAGGAAGCAAAATTCAG 6343

RESULT 22

US-09-621-976-2813/c

; Sequence 2813, Application US/09621976

; Patent No. 6639063

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621.976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 2813

; LENGTH: 832

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 235..399

US-09-621-976-2813

Query Match 3.0%; Score 45; DB 4; Length 832;

Best Local Similarity 14.3%; Pred. No. 0.057;

Matches 54; Conservative 163; Mismatches 158; Indels 2; Gaps 1;

QY 297 AACATATGTAACATAAAGATACTGGCTCAGCAAAATCAGATTCGACATGGGCTAAATC 356

Db 384 ACCATAAATAATTATATTTTCTGYTWTWKTWYWTYYTWTTRMMWKKKARRWYWMKSTY 325

QY 357 TATATCAGCTAAATACTACACAGCTTCCACTGTTACTCAGTCAAAATATATGATGGTACTGA 416

Db 324 ACASRYKCTWGMWYMKRMWSTWYCMYCMKCMYGRRCATWTTMARGRMWSYAWGKW 265

QY 417 GG--TTATTAATGTTTCCCAATATGGATATATTATTTCTAGCACTGTTAATAATGGA 474

Db 264 KNSRSAMSMCTRMYYKSGSYTWKCTCATWCYWKYWKRMWSKTCWSGSGGYMTSYT 205

QY 475 GCTGTAATATGTTTACATCTGTAACCTCAAAATGCTATTTTCATGAAAGAAATCTGAT 534

Db 204 STRSYSYMASWMTMCMWGRWNSWTYWMAGKKWRYATTTRRAMMMWMAAATMMWYM 145

QY 535 GCACATGGGGAGATGGTGGTGGGAAACTAGATCAAAACGCTACTCGGTTGGTGAT 594

Db 144 WNAWSSRSGAAYRRRTMMWGYWYWRKKSRYRTRCAWAYAWKTKRSYYWCWWMKWKRC 85

QY 595 ACAGTCAAAATACTATTACTTATAAGAAATCGAGTCAATTTATCATGGTACAGAAAAGTG 654

Db 84 MWMWMAWAGKTMMPACWKTRWRWWAWAWRWMTMMWYWYWRAMKRWMMWKWR 25

QY 655 TATCAATATGTTATAAA 671

Db 24 SWSWMMWAGWTRWAAR 8

RESULT 23

US-09-328-352-1377

; Sequence 1377, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: BAUKANNI FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328.352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 1377

; LENGTH: 11679

```

; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1377

Query Match      3.0%; Score 44.6; DB 4; Length 11679;
Best Local Similarity 44.4%; Pred. No. 0.18;
Matches 234; Conservative 0; Mismatches 284; Indels 9; Gaps 1;

QY 42 AGCATTTGGTATGGCTGTATCACCAGTTAGCGCGATAGCTTTTGGCGCTGACACAGGAC 101
DB 663 AGCATCAGGTAATGTGCACCATTAACAGTGTGTTGAAAAACATTCACGACAGATGCGGCCAA 722
QY 102 AATTACAGTTCAAGATACTCAAAAAGCGCAACTATAAAGCATATAAAGATTTTGTATGC 161
DB 723 CACGGCGTTACAGTTGTTATTAATGGGTAACTTATATGCGACTGTAGATAAAGCAGC 782
QY 162 AGAAATAGATAATGCAAAATGTATCTGATTGCAATAAAGATGAGCTTCTTATTTAATCC 221
DB 783 AGGTACATGACAGTAAGTGTACCGGCGAGTGGCTTAGTTGGGATGCAGATAAGACGAT 842
QY 222 TCAAGGTAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAACTCTTTTACGACAC 281
DB 843 TGAATGCCAAAGTAACGTTTACAGATGACGAGTAACAGCAGGTGTTAATGATACACA 902
QY 282 TACTAATGGAGGAGACATATGTAATAAAGGATGCTGCTCAGCAAAATGAGATTGC 341
DB 903 AACTTAT-----ACACTTGATACCACTGCACCAATGCACCACTGATTGACCCAGT 953
QY 342 GACATGGGCTAAATCTATATACGTAATACTACACAGTTTCCACTGTTACTGAGTCAAA 401
DB 954 TAACGGGACACAGCCGATTAACAGGTACAGCAGAACCTGGTTCAACAGTGACTGTGACTTA 1013
QY 402 TAATGATGTTACTGAGGTTAATTAATGTTTCCCAATATGATATTTATGTTTCTAGCAC 461
DB 1014 TCCGAGGCTAGTACGAAACTGTTGTAGCAGCAGCGATGTTGACTGCTGCGCAAA 1073
QY 462 TGTTAATAATGAGCTGTAAATATGTTTACATCTGTAACCTCCAAATGCTACTAATCATGA 521
DB 1074 CCCAGGTTTGAATGATGGCGATGAAGTTACAGCTGTAGCAACAGACCTGCTGGCAATAC 1133
QY 522 AAGAATACTGATGCGACATGGGAGATGGTGTGGGAAACTGTAG 568
DB 1134 ATCAGGCCCAACGACGAGTGGTGTGATGCTGTTGCGCCGACAGTAG 1180

RESULT 24
US-09-601-198-166
; Sequence 166, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 166
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-166

Query Match      2.9%; Score 44.2; DB 4; Length 606;
Best Local Similarity 44.3%; Pred. No. 0.081;
Matches 163; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 180 TGTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAGGTAAGAAGCTGA 239
DB 5826 TATTCCTAATAACAATAACAATACTATAGATTCTAATGGTATAGCTACTGTAACTCAAGG 5885
QY 240 GTATAAGCTTCAACTGATTTTAAATTTCTTTTACGACAACTACTTAATGGAGGGAGAAC 299
DB 5886 CACTCTACCAACCGGAAATATTACTGTCTAAAACCTCAATGACAAATATGTAACGTACAC 5945
QY 300 ATATGTAACTAAAAAGATAGTACGCTGAGCAAAATGAGATTGGACATGGGCTAAATCTAT 359
DB 5946 TAAACAAATAGTAGTGGATTTGCTTCAAAATACAACTGAAGATATAAGTGTTTTTCAGA 6005
QY 360 ATCAGCTAATATACACAGGATTTCCACTGTTACTGATCBAATATGATGGTACTGAGT 419
DB 6006 AAACAGTGAAGTAATGTTTACCGCTGGCATGCAAGCTAAAAATGATGGTATTAAAT 6065
```

```

Matches 181; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

QY 123 AAAAGGCGCAACCTATAAAGCATATAAAGTTTGTGTCAGAAATAGATAATGCAAAATGT 182
DB 63 AAGATATGAATTCCAAACGAAAAAACAATGCAATAAATAATTAAGTTGTGAATGA 122
QY 183 ATCTGATTTCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAGGTAAGAAGCTGAGTA 242
DB 123 TGTGATATTAGTACCTTTATATGATTTTAAAGATAAAAAATCTAACTATTATACGAAAAC 182
QY 243 TAAAGCTTCAACTGATTTAATTTCTTTTTCGACAACTACTAATGGAGGAGAACATA 302
DB 183 ATTTTATTTCCATAATTAATCAATTTGCAATGCTAAATAAAAAAATTTATTTTATTTCTAA 242
QY 303 TGTAACTAAAAAGATAGTGGCTGAGCAAAATGAGATTGCGACATGGGCTAAATCTTATATC 362
DB 243 AAAAATCTAATTAATCTGTTATTACAGAAAAATTTATTAGATATAGGATTTTACTTGAACAAT 302
QY 363 AGCTAATACTACACAGTTTCCACTGTTTACTGAGTCAAAATGAATGATGGTACTGAGTTAT 422
DB 303 TATTTATTAATGATCAAAACAAAACTTATATCAGGCCAACTGATTTTATGACCAAAAAATTT 362
QY 423 TAAATGTTTCCCAATATGATATTATTATGTTTCTAGCACTGTTAATAATGGAGCTGTAAT 482
DB 363 TATTCGATTATAAAAAAATGAATAATAATTAATCAATAGTGAATGATGAT 422
QY 483 TATGTTTACATCTGTAACTCCAAATGCTACTATTTCATGAAAAGAAATACT 531
DB 423 GATTGTTAAATGAATCAATTAATAATTAATCAATAGTGAATGATGAT 471

RESULT 25
US-09-134-001C-322
; Sequence 322, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 322
; LENGTH: 30549
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-322

Query Match      2.9%; Score 44.2; DB 4; Length 30549;
Best Local Similarity 45.2%; Pred. No. 0.32;
Matches 163; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 180 TGTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAGGTAAGAAGCTGA 239
DB 5826 TATTCCTAATAACAATAACAATACTATAGATTCTAATGGTATAGCTACTGTAACTCAAGG 5885
QY 240 GTATAAGCTTCAACTGATTTTAAATTTCTTTTACGACAACTACTTAATGGAGGGAGAAC 299
DB 5886 CACTCTACCAACCGGAAATATTACTGTCTAAAACCTCAATGACAAATATGTAACGTACAC 5945
QY 300 ATATGTAACTAAAAAGATAGTACGCTGAGCAAAATGAGATTGGACATGGGCTAAATCTAT 359
DB 5946 TAAACAAATAGTAGTGGATTTGCTTCAAAATACAACTGAAGATATAAGTGTTTTTCAGA 6005
QY 360 ATCAGCTAATATACACAGGATTTCCACTGTTACTGATCBAATATGATGGTACTGAGT 419
DB 6006 AAACAGTGAAGTAATGTTTACCGCTGGCATGCAAGCTAAAAATGATGGTATTAAAT 6065
```





RESULT 28  
US-09-889-572-3  
; Sequence 3, Application US/09889572  
; Patent No. 6610517  
; GENERAL INFORMATION:  
; APPLICANT: Lubitz, Werner  
; TITLE OF INVENTION: COMPARTMENTALIZATION OF RECOMBINANT POLYPEPTIDES IN HOST CELLS  
; FILE REFERENCE: 100564-00070  
; CURRENT APPLICATION NUMBER: US/09/889,572  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: PCT/EP00/00686  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: DE 199 03 345.5  
; PRIOR FILING DATE: 1999-01-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 3  
; LENGTH: 2766  
; TYPE: DNA  
; ORGANISM: Bacillus stearothermophilus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2763)  
; OTHER INFORMATION:  
; NAME/KEY: sig\_peptide  
; LOCATION: (1)..(93)  
; OTHER INFORMATION:  
; NAME/KEY: mat\_peptide  
; LOCATION: (94)..(2763)  
; OTHER INFORMATION:  
; US-09-889-572-3

RESULT 29  
US-09-463-402-13  
; Sequence 13, Application US/09463402  
; Patent NO. 6596510  
; GENERAL INFORMATION:  
; APPLICANT: Lubitz, Werner  
; APPLICANT: Resch, Stephanie  
; TITLE OF INVENTION: Secretion of Carrier-bound Proteins into the Periplasm  
; TITLE OF INVENTION: and into the Extracellular Space  
; FILE REFERENCE: 05649059  
; CURRENT APPLICATION NUMBER: US/09/463,402  
; CURRENT FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: DE19732829.6  
; PRIOR FILING DATE: 1997-07-30  
; PRIOR APPLICATION NUMBER: PCT/EP98/04723  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 4065  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: fusion of SbsB  
; OTHER INFORMATION: with Male  
US-09-463-402-13

RESULT 30  
US-08-257-073-10  
; Sequence 10, Application US/08257073

Patent No. 576597  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: de Taisne, Charles  
APPLICANT: Tine, John A.  
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor  
CITY: New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/257,073  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/075,783  
FILING DATE: 11-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/852,305  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,183  
FILING DATE: 20-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2570  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5181 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-257-073-10

Query Match 2.9%; Score 43.2; DB 1; Length 5181;  
Best Local Similarity 44.4%; Pred. No. 0.3;  
Matches 220; Conservative 0; Mismatches 273; Indels 3; Gaps 1;  
QY 158 ATGCAGAAATAGATAATGCAAAATGTATCTGATTGCAATAAAGATGGAGCTTCTTATTTAA 217  
Db 1130 ACGAAGAAAAATAAAGAAATGGCAAACTATTAAATTTAACATTGATGATTTATTTA 1189  
QY 218 TTCTCAAGGTAAGAAGCTGAGTATATAAGCTTCAACTGATTTAAATTCCTTTTACGA 277  
Db 1190 CTGATCCACTTCAATTAGATAATTTATTAAGAGAAAAATAAAAAAGTTGATGTAACAC 1249  
QY 278 CAACTACTAATGGAGGAGACATATGTAATAAAAAAGATACTCGCTCAGCAATGAGA 337  
Db 1250 CTAAATCACAAGATCCTACGAAATCTGTTCAAAATACCAAAAGTTCTCTTATCCAAATGGTA 1309  
QY 338 TTGCGACATGGCTAAATCTATATCAGCTAATACTACACACAGTTTCCACTGTACTGAGT 397  
Db 1310 TTG---TATATCTTTACCACTCAGTATTCATAATTCATTAGCTGCAGATAATGATA 1366  
QY 398 CAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGGATATTATTATGTTCTTA 457  
Db 1367 AAAATTCATATGGTATGATTTAATGAATCCTGTACTAAGAGAAAAAATTAATGAAAAATTA 1426  
QY 458 GCACGTGTTAATAATGGAGCTGTAATTATGGTTACATCTGTTAACTCCAAATGCTACTATTC 517

Db 1427 TTACAGATAATAAGGAAAGAAAAATATTCTATTAATAACATTAAAAAACAATGTATTAG 1486  
QY 518 ATGAAAAAGAAATACCTGATGCCGACATGGGGAGATGGTGGGAAAAAATCTGTAGATCAAAAAA 577  
Db 1487 AAGAAAAAATAATTAATCACACAAAAAGAACAAAAATTAATAAAGATTGAAAGATTATGAA 1546  
QY 578 CGTACTCGGTTGGTGATACAGTCAAAATATATCTATTCTTTAAGAATGCACTCAATTAATC 637  
Db 1547 AGTCAAAAAAGGATTATGAGAAATTTACTTGAATAATTTTATGAATAAATTAATAATA 1606  
QY 638 ATGCTACAGAAAAAGT 653  
Db 1607 ATTTGACAAAGATGT 1622

Search completed: July 27, 2004, 01:16:20  
Job time : 130 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 00:04:34 ; Search time 727 Seconds  
(without alignments)  
10146.415 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaaagaataatgattca.....gtcgtcgttaccttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3216467 seqs, 2444149694 residues

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1509	100.0	12	US-10-333-002-15	Sequence 15, Appl
2	64.8	4.3	1983	13 US-10-206-576-217	Sequence 217, Appl
3	64.8	4.3	2199	12 US-10-333-002-28	Sequence 28, Appl
4	64.8	4.3	15614	9 US-09-070-927A-45	Sequence 45, Appl
5	63.2	4.2	1881	17 US-10-661-809-12	Sequence 12, Appl
6	59.2	3.9	1137	13 US-10-282-122A-15988	Sequence 15988, A
7	53	3.5	4985	15 US-10-094-240-10	Sequence 10, Appl
8	53	3.5	4985	15 US-10-056-405-10	Sequence 10, Appl
9	50.4	3.3	14066	15 US-10-349-680-149	Sequence 149, Appl
10	50.4	3.3	14067	13 US-10-282-122A-40681	Sequence 40681, A
11	50	3.3	3996	15 US-10-087-464-42	Sequence 42, Appl
12	49.8	3.3	1687	13 US-10-206-576-219	Sequence 219, Appl
13	48.6	3.2	4997	13 US-10-282-122A-35506	Sequence 35506, A
14	48.4	3.2	3666	9 US-09-137-531-13	Sequence 13, Appl

48.4	3.2	3666	9	US-09-137-531-14	Sequence 14, Appl
48.4	3.2	4197	9	US-09-137-531-7	Sequence 7, Appl
48.4	3.2	4197	9	US-09-137-531-8	Sequence 8, Appl
47.4	3.1	1037	16	US-10-373-667-3	Sequence 3, Appl
47.4	3.1	1159	16	US-10-373-667-1	Sequence 1, Appl
47.4	3.1	1471	16	US-10-373-667-2	Sequence 2, Appl
47.4	3.1	4047	9	US-09-815-242-4843	Sequence 4843, Ap
47.4	3.1	4050	9	US-09-815-242-9039	Sequence 9039, Ap
47.4	3.1	4358	8	US-08-781-986A-454	Sequence 454, App
47.4	3.1	4358	13	US-10-329-624-454	Sequence 1, Appl
47.2	3.1	5314	15	US-10-155-533-1	Sequence 1, Appl
46.8	3.1	2801	13	US-10-282-122A-34844	Sequence 34844, A
46.8	3.1	7047	15	US-10-240-453-260	Sequence 260, App
46.2	3.1	4158	13	US-10-282-122A-8140	Sequence 8140, Ap
45.8	3.0	684707	16	US-10-398-221-9	Sequence 9, Appl
45.8	3.0	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
45.6	3.0	2017	15	US-10-155-533-3	Sequence 3, Appl
45.6	3.0	6397	13	US-10-221-714A-181	Sequence 181, App
45.6	3.0	6397	15	US-10-239-676-107	Sequence 107, App
45.6	3.0	6397	15	US-10-311-455-1319	Sequence 1319, Ap
45.6	3.0	6397	15	US-10-240-453-119	Sequence 119, App
45.4	3.0	2217	13	US-10-282-122A-17439	Sequence 17439, A
45.2	3.0	3931	15	US-10-006-780-1	Sequence 1, Appl
45	3.0	12237	15	US-10-311-455-2331	Sequence 2331, Ap
44.8	3.0	573	15	US-10-029-386-25433	Sequence 25433, A
44.8	3.0	5358	17	US-10-637-544-19	Sequence 19, Appl
44.8	3.0	16287	15	US-10-311-455-645	Sequence 645, App
44.4	2.9	7446	13	US-10-282-122A-7578	Sequence 7578, Ap
44.2	2.9	486	13	US-10-465-217-13	Sequence 13, Appl
44.2	2.9	606	15	US-10-349-680-85	Sequence 85, Appl
44	2.9	3673778	15	US-10-313-841-1	Sequence 1, Appl
43.8	2.9	528	15	US-10-029-386-5350	Sequence 5350, Ap
43.8	2.9	807	13	US-10-282-122A-16974	Sequence 16974, A
43.8	2.9	1243	13	US-10-282-122A-9076	Sequence 9076, Ap
43.8	2.9	1910	15	US-10-155-533-2	Sequence 2, Appl
43.8	2.9	3532	15	US-10-017-161-1913	Sequence 1913, Ap
43.8	2.9	3532	16	US-10-292-798-1569	Sequence 1569, Ap
43.6	2.9	1590	15	US-10-032-585-6364	Sequence 6364, Ap
43.4	2.9	2000	9	US-09-938-842A-3954	Sequence 3954, Ap
43.4	2.9	2000	11	US-09-938-842A-3954	Sequence 3954, Ap
43.2	2.9	475	9	US-09-864-761-1619	Sequence 1619, Ap
43	2.8	483	16	US-10-398-221-365	Sequence 365, App
43	2.8	483	16	US-10-398-221-2558	Sequence 2558, Ap
43	2.8	2001	9	US-09-737-178-84	Sequence 84, Appl
43	2.8	2001	10	US-09-853-079-84	Sequence 84, Appl
43	2.8	2001	17	US-10-294-443-84	Sequence 84, Appl
43	2.8	3057	15	US-10-349-680-148	Sequence 148, App
43	2.8	3402	9	US-09-737-178-86	Sequence 86, Appl
43	2.8	3402	10	US-09-853-079-86	Sequence 86, Appl
43	2.8	3402	17	US-10-294-443-86	Sequence 86, Appl
43	2.8	8201	15	US-10-363-798-1	Sequence 1, Appl
42.8	2.8	1959	9	US-09-864-761-4012	Sequence 4012, Ap
42.8	2.8	2058	13	US-10-282-122A-7579	Sequence 7579, Ap
42.8	2.8	2769	15	US-10-032-585-6675	Sequence 6675, Ap
42.6	2.8	663	13	US-10-027-632-250796	Sequence 250796, A
42.6	2.8	663	16	US-10-027-632-250796	Sequence 250796, A
42.6	2.8	663	16	US-10-027-632-250796	Sequence 250796, A
42.6	2.8	663	16	US-10-027-632-250797	Sequence 250797, A
42.6	2.8	2247	13	US-10-424-559-44419	Sequence 44419, A
42.6	2.8	168575	15	US-10-178-194-1	Sequence 1, Appl
42.4	2.8	1224	13	US-10-282-122A-16026	Sequence 16026, A
42.4	2.8	1495	16	US-10-398-221-3257	Sequence 3257, Ap
42.4	2.8	2766	9	US-09-117-447-5	Sequence 5, Appl
42.4	2.8	5917	15	US-10-087-464-9	Sequence 9, Appl
42.2	2.8	810	15	US-10-349-680-126	Sequence 126, App
42.2	2.8	3426	13	US-10-282-122A-34067	Sequence 34067, A
42.2	2.8	4185	15	US-10-304-095-7	Sequence 7, Appl
42.2	2.8	10640	15	US-10-304-095-5	Sequence 5, Appl
42.2	2.8	58985	10	US-09-901-152-3	Sequence 3, Appl
42.2	2.8	134601	10	US-09-855-824-3	Sequence 3, Appl
42	2.8	5379	15	US-10-311-455-1650	Sequence 1650, App
42	2.8	5379	15	US-10-240-485-130	Sequence 130, App
42	2.8	6060	8	US-08-781-986A-534	Sequence 534, App

c	88	42	2.8	6050	13	US-10-329-624-534	Sequence 534, App
	89	42	2.8	7104	9	US-09-815-242-4580	Sequence 4580, Ap
	90	42	2.8	7107	9	US-09-815-242-8291	Sequence 8291, Ap
ALIGNMENTS							
RESULT 1							
US-10-333-002-15							
; Sequence 15, Application US/10333002							
; Publication No. US20040071729A1							
; GENERAL INFORMATION:							
; APPLICANT: Adderson, Elisabeth							
; APPLICANT: Bohnsack, John							
; TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic							
; FILE OF INVENTION: Compositions and Vaccines Thereof							
; FILE REFERENCE: 2511-1-001 (S8-0039)							
; CURRENT APPLICATION NUMBER: US/10/333,002							
; CURRENT FILING DATE: 2003-07-08							
; PRIOR FILING DATE: 2001-07-08							
; PRIOR FILING DATE: 2001-08-08							
; PRIOR APPLICATION NUMBER: PCT/US01/24795							
; PRIOR APPLICATION NUMBER: US 09/634,341							
; PRIOR FILING DATE: 2000-08-08							
; NUMBER OF SEQ ID NOS: 37							
; SOFTWARE: PatentIn version 3.1							
; SEQ ID NO 15							
; LENGTH: 1509							
; TYPE: DNA							
; ORGANISM: Streptococcus agalactiae							
US-10-333-002-15							
Query Match							
Best Local Similarity 100.0%; Score 1509; DB 12; Length 1509;							
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
QY	1	ATGAAAAAGAAATGATTC	CAATCGCTGTTAGTGGCGAGTTT	AGCAATTCGTTATGGCTGTA	60		
DB	1	ATGAAAAAGAAATGATTC	CAATCGCTGTTAGTGGCGAGTTT	AGCAATTCGTTATGGCTGTA	60		
QY	61	TCACGAGTTACGCGATAG	CTTTGCGCTGAGACAGGCAAT	TACAGTTCAAGTACT	120		
DB	61	TCACGAGTTACGCGATAG	CTTTGCGCTGAGACAGGCAAT	TACAGTTCAAGTACT	120		
QY	121	CAAAAAGGCGCAACCTA	TAAAGCATATAAAGTTT	TGATGCGAAATAGATAATGCAAT	180		
DB	121	CAAAAAGGCGCAACCTA	TAAAGCATATAAAGTTT	TGATGCGAAATAGATAATGCAAT	180		
QY	181	GTATCTGATTGGAATAA	GATGAGGCTTCTTATTT	TAATTCCTCAAGGTAAGAAGCTGAG	240		
DB	181	GTATCTGATTGGAATAA	GATGAGGCTTCTTATTT	TAATTCCTCAAGGTAAGAAGCTGAG	240		
QY	241	TATAAGGCTTCACTGAT	TTTAATTCCTTTTACGCAACTACT	TAATGGAGGAGAAC	300		
DB	241	TATAAGGCTTCACTGAT	TTTAATTCCTTTTACGCAACTACT	TAATGGAGGAGAAC	300		
QY	301	TATGTAATCAAAAAGAT	ACTGGCTCAGCAAAATGAG	ATTGCGACATGGGCTAAATCTATA	360		
DB	301	TATGTAATCAAAAAGAT	ACTGGCTCAGCAAAATGAG	ATTGCGACATGGGCTAAATCTATA	360		
QY	361	TCAGCTAACTACACAG	TTTCCACTGTTACTGAGTCAAA	TAAATGATGGTACTGAGGTT	420		
DB	361	TCAGCTAACTACACAG	TTTCCACTGTTACTGAGTCAAA	TAAATGATGGTACTGAGGTT	420		
QY	421	ATTAAATGTTCCCAAT	ATGATTAATTTCTTAGCAGCTGTT	TAATGGAGCTGTA	480		
DB	421	ATTAAATGTTCCCAAT	ATGATTAATTTCTTAGCAGCTGTT	TAATGGAGCTGTA	480		
QY	481	ATTATGTTTACATCTG	TAACTCCAAATGCTACTATT	TTCATGAAAGAAATACTGATGCGACA	540		
DB	481	ATTATGTTTACATCTG	TAACTCCAAATGCTACTATT	TTCATGAAAGAAATACTGATGCGACA	540		
QY	541	TGGGGAGATGGTGGT	GGAACAACTAGATCAAAAAAC	GTACTCGCTTGGTGATACAGTC	600		

541	TGGGGAGATGGTGGTGG	AAAAACGTAGATCAAAAAAC	GTACTCGGTTGGTGATACAGTC	600			
601	AAATATACTATTACTAT	AAGAAATGCAATTCATCAT	GTGATGAGAAAAAGTGTATCAA	660			
601	AAATATACTATTACTAT	AAGAAATGCAATTCATCAT	GTGATGAGAAAAAGTGTATCAA	660			
661	TATGTTATAAAGGATAC	TATGCCATCTCTCTGCTAG	TTGAAACGAAGGTCCTTAT	720			
661	TATGTTATAAAGGATAC	TATGCCATCTCTCTGCTAG	TTGAAACGAAGGTCCTTAT	720			
721	GAAATTAATCTATTTAC	TGATGGATCAAGGAAATTA	CAACTCAAGGTTTCGAAAAA	780			
721	GAAATTAATCTATTTAC	TGATGGATCAAGGAAATTA	CAACTCAAGGTTTCGAAAAA	780			
781	GCACTGGGAAGTATAAC	CTGTTAGAGGAAATAATA	TATTTCAAGATTAATTCGCTGG	840			
781	GCACTGGGAAGTATAAC	CTGTTAGAGGAAATAATA	TATTTCAAGATTAATTCGCTGG	840			
841	GCAGCTACCAATACTCC	CAACCGGAAATACTCAAA	TGGAGCTTAATGATGACTTTT	900			
841	GCAGCTACCAATACTCC	CAACCGGAAATACTCAAA	TGGAGCTTAATGATGACTTTT	900			
901	AAGGGAATAAATAACA	TACCAATCACAGTCACTT	TATACAGGATTAATAAGAGTGA	960			
901	AAGGGAATAAATAACA	TACCAATCACAGTCACTT	TATACAGGATTAATAAGAGTGA	960			
961	GGTTCAGCTGATTTAC	CAGAAAAATACAAATTC	GGACCAATCAACCCCAATTA	1020			
961	GGTTCAGCTGATTTAC	CAGAAAAATACAAATTC	GGACCAATCAACCCCAATTA	1020			
1021	GATGACCCAGGTCAAAA	AGTAAACAGTGGGATGCT	CAAAATTAATTAATAAAAAAT	1080			
1021	GATGACCCAGGTCAAAA	AGTAAACAGTGGGATGCT	CAAAATTAATTAATAAAAAAT	1080			
1081	GGTTCACAAAGCTTCTA	TACAGGCTATATTTGTTT	TAAAGAAATGCTACGGGTCAA	1140			
1081	GGTTCACAAAGCTTCTA	TACAGGCTATATTTGTTT	TAAAGAAATGCTACGGGTCAA	1140			
1141	TTTCTTAACTTTTAA	CGATACAAATTAACGTTG	GAATGGGCAACAGAGCTAA	1200			
1141	TTTCTTAACTTTTAA	CGATACAAATTAACGTTG	GAATGGGCAACAGAGCTAA	1200			
1201	TATACACAGGAGCAGAT	GGTATTAATTAATTAATTA	CAAGCTTTGAAAGAGGTA	1260			
1201	TATACACAGGAGCAGAT	GGTATTAATTAATTAATTA	CAAGCTTTGAAAGAGGTA	1260			
1261	CTAGTTGAGAAAAAGG	CTCCCTTAGGTTACAAAT	TTTGTAGATACTCTCAGAGGTTAT	1320			
1261	CTAGTTGAGAAAAAGG	CTCCCTTAGGTTACAAAT	TTTGTAGATACTCTCAGAGGTTAT	1320			
1321	TTAGGAGATGGAGCC	CACTGATACAAATTAAC	CTTTTAGTTAAACCAACTGTT	1380			
1321	TTAGGAGATGGAGCC	CACTGATACAAATTAAC	CTTTTAGTTAAACCAACTGTT	1380			
1381	GAAAAATAAAGGTA	CTGAGTTGCTTTCAAC	CAGGTGGTATTTGGTACAACTTTCTAC	1440			
1381	GAAAAATAAAGGTA	CTGAGTTGCTTTCAAC	CAGGTGGTATTTGGTACAACTTTCTAC	1440			
1441	ATTATAGTGCATTTT	TAGTAATAGGACAGGTA	TCGTTGCTTGTCTCGTCTGTTTA	1500			
1441	ATTATAGTGCATTTT	TAGTAATAGGACAGGTA	TCGTTGCTTGTCTCGTCTGTTTA	1500			
1501	CGTTCCTTAA	1509					
1501	CGTTCCTTAA	1509					

RESULT 2  
US-10-206-576-217  
; Sequence 217, Application US/10206576  
; Publication No. US20030017495A1  
; GENERAL INFORMATION:

APPLICANT: Choi et al.  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 497  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-R  
COMPUTER: Dell Latitude  
OPERATING SYSTEM: Windows 98  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/206,576  
FILING DATE: 29-Jul-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/071,035  
FILING DATE: 1998-05-04  
APPLICATION NUMBER: US 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: US 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: US 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Hyman, Mark J.  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB369PID1  
INFORMATION FOR SEQ ID NO: 217:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1983 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 217:  
US-10-206-576-217  
Query Match 4.3%; Score 64.8; DB 13; Length 1983;  
Best Local Similarity 52.1%; Pred. No. 0.0012;  
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;  
QY 1180 ACAGAGCTAATGCAACAGAAATATACACAGGAGCAGATGGTATTAATTACCAATTACAGGC 1239  
DB 1660 ACAAAGCTGAAGCAACTACTTTTACACACACGGCTGATGGATTAGTTGATATACAGGG 1719  
QY 1240 TTGAAGAAGGTACATACATCTAGTTGAGAAAAAGGCTCCCTTAGGTTACAAATTTGTTA 1299  
DB 1720 CTTAAATACGGTACCTATTATTAGAAAGAACTGTAGCTCCTGATGATTTATGCTTTGTTA 1779  
QY 1300 GATAACTCTCAGAGGTTATTTTAGGAGATGAGCCACTGATACCACTAAATTCAGATAAC 1359  
DB 1780 ACAATCGGATGAATTTGTGTCATGCAATCAATATGACAA-----CAGAAAC 1833  
QY 1360 CTTTGTAGTAAACCACTGCGTAAATACAAAGGTACTGAGTTGCTTCAACAGGTGTT 1419  
DB 1834 CTAGTTTACCAAGAAAAGTACCAACAAACAAAGGTACCTTACCTTCAACAGGTGGC 1893  
QY 1420 ATTGGTACAAATTTTCTACATTAATAGTGCAATTTTAGTAAATAGGACAGGTATCGTG 1479  
DB 1894 AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTATTATTCAGGAGTCTAC 1953  
QY 1480 CTTGTGCTCGTCGTTTACGTTCTTAA 1509  
DB 1954 TTGCTAGACGTAGAAAAGAAATGCTTAA 1983  
RESULT 3  
US-10-333-002-28  
; Sequence 28, Application US/10333002

Publication No. US20040071729A1  
GENERAL INFORMATION:  
APPLICANT: Adderson, Elisabeth  
APPLICANT: Bohnsack, John  
TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic  
FILE OF INVENTION: Compositions and Vaccines Thereof  
FILE REFERENCE: 2511-1-001 (S7-0039)  
CURRENT APPLICATION NUMBER: US/10/333,002  
CURRENT FILING DATE: 2003-07-08  
PRIOR APPLICATION NUMBER: PCT/US01/24795  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 09/634,341  
PRIOR FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 28  
LENGTH: 2199  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-10-333-002-28  
Query Match 4.3%; Score 64.8; DB 12; Length 2199;  
Best Local Similarity 52.1%; Pred. No. 0.0013;  
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;  
QY 1180 ACAGAGCTAATGCAACAGAAATATACACAGGAGCAGATGGTATTAATTACCAATTACAGGC 1239  
DB 275 ACAAAGCTGAAGCAACTACTTTTACAAACACGGCTGATGGATTAGTTGATATACAGGG 334  
QY 1240 TTGAAGAAGGTACATACATCTAGTTGAGAAAAAGGCTCCCTTAGGTTACAAATTTGTTA 1299  
DB 335 CTTAAATACGGTACCTATTATTAGAAAGAACTGTAGCTCCTGATGATTTATGCTTTGTTA 394  
QY 1300 GATAACTCTCAGAGGTTATTTTAGGAGATGAGCCACTGATACCACTAAATTCAGATAAC 1359  
DB 395 ACAATCGGATGAATTTGTGTCATGCAATCAATATGACAA-----CAGAAAC 448  
QY 1360 CTTTGTAGTAAACCACTGTTGAGAAATACAAAGGTACTGAGTTGCTTCAACAGGTGTT 1419  
DB 449 CTAGTTTCAACGAAAAAGTACCAACAAACAAAGGTACTTACCTTCAACAGGTGGC 508  
QY 1420 ATTGGTACAAATTTTCTACATTAATAGTGCAATTTTAGTAAATAGGACAGGTATCGTG 1479  
DB 509 AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTATTATTCAGGAGTCTAC 568  
QY 1480 CTTGTGCTCGTCGTTTACGTTCTTAA 1509  
DB 569 TTGCTAGACGTAGAAAAGAAATGCTTAA 598  
RESULT 4  
US-09-070-927A-45  
; Sequence 45, Application US/09070927A  
; Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunsch  
Patrick J. Dillon  
Steven Barash  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:



```
/ APPLICATION NUMBER: US/09/070,927A
/ FILING DATE: 04-May-2000
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/046,655
/ FILING DATE: 1997-05-16
/ APPLICATION NUMBER: 60/044,031
/ FILING DATE: 1997-05-06
/ APPLICATION NUMBER: 60/066,009
/ FILING DATE: 1997-11-14
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kenley K. Hoover
/ REGISTRATION NUMBER: 40,302
/ REFERENCE/DOCKET NUMBER: PB369
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 45:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 15614 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-070-927A-45

Query Match          4.3%; Score 64.8; DB 9; Length 15614;
Best Local Similarity 52.1%; Pred. No. 0.0034;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAAAGCTAATGCAACAGAAATATACAAACAGGAGCAGATGGTATATACCAATTACAGGC 1239
Db 13724 ACAAAGCTGAAGCAACTACTTTTACAAACAGCGCTGATGGATTAGTTGATATACAGGC 13783

QY 1240 TTGAAAGAGGTACATCTACTCTAGTTGAGAAAAGCGTCCCTTAGGTTACAATTGTTA 1299
Db 13784 CTTAAATACGGTACCTATTATTATTAGAGAAACTGTAGCTCCCTGATGATTATGCTTTGTTA 13843

QY 1300 GATAACTCTCAGAAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 13844 ACAATCGGATTGAATTTGGTCAATGAACATCATATGGCACAA-----CAGAAAC 13897

QY 1360 CTTTGTAGTTAACCAACTGTTGAAAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 13898 CTAGTTTCCAGAAAAGTACCAAAACAAACAAAGGTACCTTCAACAGGTGGC 13957

QY 1420 ATTGGTACAACTTTCTACATTTAGTGCAATTTTAGTATAGGACGAGTATCGTG 1479
Db 13958 AAAGGAATCTACGTTTACTTAGGAAGTGGCGCAGTCTTGCTACTTATTGACGAGTCTAC 14017

QY 1480 CTTGTTGCTCGTGGTCTTTACGTTCTTAA 1509
Db 14018 TTTGCTAGAGTGAAGAAAGAAATGCTTAA 14047

RESULT 5
US-10-661-809-12
; Sequence 12, Application US/10661809
; Publication No. US2004010191A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR
; FILE REFERENCE: P07741US01/BAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1881
; TYPE: DNA

/ ORGANISM: Staphylococcus epidermidis
US-10-661-809-12

Query Match          4.2%; Score 63.2; DB 17; Length 1881;
Best Local Similarity 52.5%; Pred. No. 0.0026;
Matches 165; Conservative 0; Mismatches 143; Indels 6; Gaps 1;

QY 1180 ACAGAAAGCTAATGCAACAGAAATATACAAACAGGAGCAGATGGTATATACCAATTACAGGC 1239
Db 1561 ACAAAGCTGAAGCAACTACTTTTACAAACAGCGGTGATGGATTAGTTGATATACAGGC 1620

QY 1240 TTGAAAGAGGTACATCTACTCTAGTTGAGAAAAGGCTCCCTTAGGTTACAATTGTTA 1299
Db 1621 CTTAAATACGGTACCTATTATTATTAGAGAAACTGTAGCTCCCTGATGATTATGCTTTGTTA 1680

QY 1300 GATAACTCTCAGAAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 1681 ACAATCGGATTGAATTTGTGTCATGAATCATATGGCACAA-----CAGAAAC 1734

QY 1360 CTTTGTAGTTAACCCAACTGTTGAAAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 1735 CTAGTTTCCAGAAAAGTACCAAAACAAACAAAGGTACCTTCAACAGGTGGC 1794

QY 1420 ATTGGTACAACTTTTCTACATTTAGTGCAATTTTAGTATAGGACGAGTATCGTG 1479
Db 1795 AAAGGAATCTACGTTTACTTAGGAAGTGGCGCAGTCTTGCTACTTATTGACGAGTCTAC 1854

QY 1480 CTTGTTGCTCGTGG 1493
Db 1855 TTTGCTAGACGTAG 1868

RESULT 6
US-10-282-122A-15988
; Sequence 15988, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

```

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15988
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15988

```

Query Match	3.9%;	Score 59.2;	DB 13;	Length 1137;
Best Local Similarity	43.7%;	Pred. No. 0.015;		
Matches 262;	Conservative 0;	Mismatches 338;	Indels 0;	Gaps 0;
QY	90	TGAGACAGGACAAATTACAGTTCAAGATACCTCAAAAAGGCGCAACCTATAAAGCATATAA 149		
Db	30	TGGGACAAGCCGAGAGCTATAAGATGGCTCTCTTTGGTTAAAAAACTGAAAAGAAATAA 89		
QY	150	AGTTTTTGATGCGAGAAATAGATAATGCAAAATGTATCTGATTCGAATAAAGATGGAGCTTC 209		
Db	90	AAATTTTTCAAGCAAAAGTATGTGTTACAGCACAGCACAGGGAGATGCTAGATCAAGTTCT 149		
QY	210	TTATTTTAA'TCCCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAA'TTCTCT 269		
Db	150	TAGTTTGTTCAAATATAGCCAGCATTTGATTTGAATATAATGAAAATAAGCAATCTCT 209		
QY	270	TTTTTACGACAACTACTAATGAGGGAGAAACATATGTAACTAAAAAAGATACTGGCTCAGC 329		
Db	210	TACAACGATAACTTCAAGTGTTTTACATGGGTTAAGCGAATATTTCTGCTGAAAGACC 269		
QY	330	AAATGAGATTGCGGACATGGGCTAAATCTATATCAGCTAATACTACACAGTTTCCACGTG 389		
Db	270	AGATATATTTTGTAGATGATGAGATACAACTACACACGTTTGCAGCATCACTTGCAGCCTT 329		
QY	390	TACTGAGTCAAAATAATGATGCTAGTGAGTTTATTAATGTTTTCCCAATATGGATATTA 449		
Db	330	TTATGAAAAGATAGCTATATGGGCATTTGAGCTGGACCTTAGAACTTATGATATAATATTT 389		
QY	450	TGTTTTCTAGCACCTGTAATAATGGAGCTGTAA'TTATGGTTACATCTGTAACTCCAATGCG 509		
Db	390	TCCTTTTTCTGAGCAAGTAAATAGAAAACCTTACTGGAGCAATTCAGATATGCATTTTGC 449		
QY	510	TACTATTTCTGAAAAGAAATCTGATGCGACATGGGGAGATGGTGGTGGAAAAA'CTGTAGA 569		
Db	450	ACCGACAGTAAAATCTAAAAATAATCTTTTAAAGAGAAAGGCTGTAAAAGAAAAAATATATT 509		
QY	570	TCAAAAAACGTA'CTCCGGTTGGTGATACAGTCAAAATATACTATTTACTTTATAAGATGCAGT 629		
Db	510	TATAA'CTGGAAAATACTGTGATTGATGCTATGAAATACACAGTGGATAGTAA'TTATGTATT 569		
QY	630	CAATTTATCATGGGTACAGAAAAAGTGTATCAATATGTTTATAAAGGATACTATGCCATCTGC 689		
Db	570	TAAAAATGATCAATTTAAATAAATTTGGACTACAAACATAAAAAGAGTTATTA'TGGTAACTGC 629		

## RESULT 7

```

US-10-094-240-10/c
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
;
; GENERAL INFORMATION:
; APPLICANT: ZWIBEL, LAURENCE J.
; TITLE OF INVENTION: ARRESTIN GENE,
; FILE REFERENCE: N9289
; CURRENT APPLICATION NUMBER: US/10/0
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 10/056,40
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,64
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent In ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
;

```

; ORGANISM: Anopheles gambiae  
US-10-094-240-10

Query Match 3.5%; Score 53; DB 15; Length 4985;  
Best Local Similarity 44.9%; Pred. No. 0.66;  
Matches 241; Conservative 0; Mismatches 295; Indels 1; Gaps 1;

Qy 170 ATAATGCAAAATGATCTGATTCGAATAAAGATGGAGCTCTCTATTTTAAATTCCTCAAGTA 229  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2854 ATAATAATATTAATAATAATAATATTAATAATAATAATAATAATAATAATAATAATAATA 2795  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 230 AAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATCTCTTTTTTACGACAACACTACTAATG 289  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2794 ATA 2735  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 290 GAGGAGAACATATGTAACTAAAAAGATACCTGGTCAGCAAAATGAGATGCGCACATGGG 349  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2734 TGATAATAATGTATAATAATGAATAACAATAATAATAATAATAATAATAATAATAATAA 2675  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 350 CTAATCTATATCA-GCTAATATCTACACCAGTTTCACACTGTTACTGAGTCAAAATAATGAT 408  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2674 TTAA 2615  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 409 GGTACTGAGGTATTAATGTTTCCCAATATGGATATTAATGTTTCTAGCACCTGTTAAT 468  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2614 AATAATAATAATCATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAA 2555  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 469 AATGGAGCTGTAATATATGTTTACATCTGTAACCTCCAAATGCTACTATTCATGAAAGAAT 528  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2554 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 2495  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 529 ACTGATCGACATGGGGAGATGGTGGGAAAAAACTGTAGATCAAAAAACGTACTCGGTT 588  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2494 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 2435  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 589 GGTGATACAGTCAAAATATCTATTACTTTAAGAAGTCAGTCAATTATCATGGTACAGAA 648  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2434 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 2375  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 649 AAAAGTGATCAATATGTTATPAAAGGATCATGCCATCTGCTGTGAGTGGATTG 705  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2374 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 2318  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8  
US-10-056-405-10/c  
; Sequence 10, Application US/10056405  
; Publication No. US20030166013A1  
; GENERAL INFORMATION:  
; APPLICANT: ZWIEBEL, LAURENCE J.  
; TITLE OF INVENTION: MOSQUITO OLFACTORY GENES, POLYPEPTIDES, AND METHODS OF  
; TITLE OF INVENTION: USE THEREOF  
; FILE REFERENCE: N7841  
; CURRENT APPLICATION NUMBER: US/10/056,405  
; CURRENT FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: 60/264,649  
; PRIOR FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 4985  
; TYPE: DNA  
; ORGANISM: Anopheles gambiae  
US-10-056-405-10

## RESULT 8

```

US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFACTORY GENES, POLYPEPTIDES, AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

Query Match      3.5%; Score 53; DB 15; Length 4985;
Best Local Similarity 44.9%; Pred. No. 0.66;
Matches 241; Conservative 0; Mismatches 295; Indels 1; Gaps

Qy      170 ATAAATGCAATGTATCTGATTGCAATAAGATGGAGCTCTTATTATTATTCCTCAAGGTA 22
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2854 ATAAATATATTTAAAAATAATATTAATAATAATAATAATAATAATAATAATAATAATA 27
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

QY 230 AAGAAGCTGAGTATAAGCTTCAACTGATTTTAAATCTCTTTTACGACAACTACTAATG 289  
DB 2794 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2735  
QY 290 GAGGAGACATATGTAATCACTTAAAGAGATACCTGGCTCAGCAATGAGATTGGACATGGG 349  
DB 2734 TGATAATAATGTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2675  
QY 350 CTAAATCTATATCA-GCTAATACTACACCACTGTTTCCACTGTTTCTGAGTCAAAATATGAT 408  
DB 2674 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2615  
QY 409 GGTACTGAGGTATTAATGTTTCCCAATATGATGATATTAATGTTTCTAGCACTGTTAAT 468  
DB 2614 AATAATAATAATCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2555  
QY 469 AATGGAGCTGAATTAATGTTTACATCTGTAACCTCCAAATGCTACTATTCATGAAAGAA 528  
DB 2554 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2495  
QY 529 ACTGATCGCATGGGAGATGGTGGGAAACTGTAGATCAAAAACGTACTCGGTT 588  
DB 2494 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2435  
QY 589 GGTGATACAGTCAAAATATCTACTTATAAGAAATGCAATCAATATCATGTCACAGAA 648  
DB 2434 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2375  
QY 649 AAAGTGTATCAATATGTTTAAAGGATACATATGCCATCTGCTTCTGTAGTTGATTTG 705  
DB 2374 AATAATAATAATAATAATAATAATAATAATAATAATCTGTATTTGGTTCCTGTAACAAATG 2318

RESULT 9  
US-10-349-680-149/c  
; Sequence 149, Application US/10349680  
; Publication No. US20030176654A1  
; GENERAL INFORMATION:  
; APPLICANT: Cassell, Gail  
; APPLICANT: Chen, Ellison  
; APPLICANT: Glass, Jennifer  
; APPLICANT: Glass, John  
; APPLICANT: Heiner, Cheryl  
; APPLICANT: Lefkowitz, Elliot  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA  
; TITLE OF INVENTION: UREA LYTICUM  
; FILE REFERENCE: UAB-13403/22  
; CURRENT APPLICATION NUMBER: US/10/349,680  
; CURRENT FILING DATE: 2003-01-23  
; PRIOR APPLICATION NUMBER: US 09/601,198  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: PCT/US99/01972  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: US 60/073,189  
; PRIOR FILING DATE: 1998-01-30  
; NUMBER OF SEQ ID NOS: 181  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 149  
; LENGTH: 14066  
; TYPE: DNA  
; ORGANISM: Ureaplasma urealyticum  
US-10-349-680-149

Query Match 3.3%; Score 50.4; DB 15; Length 14066;  
Best Local Similarity 42.8%; Pred. No. 4;  
Matches 313; Conservative 0; Mismatches 416; Indels 3; Gaps 1;  
QY 435 ATATGATATATATATGTTTCTAGCACTGTTTAAATGAGCTGTAATATGTTTACATC 494  
DB 5284 AAATGTTAGTTCTTAATGATTTCTATAAATACTACACACAAACAGTTAATATGATATGA 5225  
QY 495 TGTAACTCAATGCTACTATTTCATGAAAAGAAATCTATGCGACATGGGAGATGGTGG 554

DB 5224 TGGAAATTCACGAACTTGAATTAATAAAAAAATTCATTTGGTATATATCTTCAAAATGACGG 5165  
QY 555 TGGAAAACCTGTAGATCAAAAAACGTTACTCGGTGGTGATACAGTCAAAATATATCTATTAC 614  
DB 5164 TGAAGAAATTTTAAGTGATCAAAAAACATTCGCTTTGGGCTAATAACACTATATGTTTCGA 5105  
QY 615 TTATAAGAAATCGAGTCAATTTATCATGTCAGAAAAAGTGTATCAATATGTTTATAAGGA 674  
DB 5104 ATTGCTCAATTTTAAACATTAATCGTAAATATACATAAAGAAAGTAAAGATTTAATGA 5045  
QY 675 TACTATGCCATCTGCTTCTGTAGTTGATTTGAAGAAAGGTCCTTATGAAGTACTATTAC 734  
DB 5044 TGATAATAAACATCAATAATTTTCCATTTTAAAAAATGGAATAGCTGATTTATTTGT 4985  
QY 735 TGATGGATCAGGAATATTACAACTCTAACTCAAGGTTTCGAAAAAAGCAACTGGGAAGTA 794  
DB 4984 TAAATAAACACATCGAATAGTATTAAGTTCAATTAATTGAGCCAAATTGCAAGAGCGAAAA 4925  
QY 795 TAACTCTGTTAGAGGAAAAATAATTTTCAGATTTACTATTCGGTGGGCGAGCTACCAATAC 854  
DB 4924 TAAATTTACAATCAACACAAATTAGATTTTATATTAATTAATGATCTCTGATTAATGTTTAAAGTAA 4865  
QY 855 TCCAACCGGAATACTCAAAATGAGCTAATGATGACTTTTTTTTATAAGGGAATAAATAC 914  
DB 4864 TGAAGAAAGCAATCAATTAATTTATGTTGAGAAAT---TTATCTGTAAGCAAAAGTTAA 4808  
QY 915 AATCAGCTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACAGGTTTCAGCTGATTT 974  
DB 4807 AATTGTAGGAGTCAAAAAATTTTAGAAACAACTCAATTAATCTAGCAATTAACCAAGA 4748  
QY 975 ACCAGAAATCAACATTTGGCCATCAACCCCAATCTAGCAATGATGACCCAGGTCA 1034  
DB 4747 TACTATTATTAACAGTATTTACTTTTATTAATAAACCAATAAAAGCTGCTAGCAATTTGG 4688  
QY 1035 AAAAGTAACTGAGGGATGTCATAATTAATAAAAAAATTTGATGTTTCCACAAAAGC 1094  
DB 4687 AATAGATAATGACATAAATAATATATACAAATACAAATAATAATCCATTAATAAATGTA 4628  
QY 1095 TTCATTACAAGGTGCTATATTTGTTTAAAGAAATGCTACGGTCAATTTCTTAACATTAA 1154  
DB 4627 TAATAATTTTAGTGTCAATTTGGTCTCTGTTTCAAAAGTGATACTAATAAACGATATTACTTTTGGGA 4568  
QY 1155 CGATCAATAA 1166  
DB 4567 ATATAATAATAA 4556

RESULT 10  
US-10-282-122A-40681  
; Sequence 40681, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangseu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26		
; PRIOR APPLICATION NUMBER: 60/230,335		
; PRIOR FILING DATE: 2000-09-06		
; PRIOR APPLICATION NUMBER: 60/230,347		
; PRIOR FILING DATE: 2000-09-09		
; PRIOR APPLICATION NUMBER: 60/242,578		
; PRIOR FILING DATE: 2000-10-23		
; PRIOR APPLICATION NUMBER: 60/253,625		
; PRIOR FILING DATE: 2000-11-27		
; PRIOR APPLICATION NUMBER: 60/257,931		
; PRIOR FILING DATE: 2000-12-22		
; PRIOR APPLICATION NUMBER: 60/267,636		
; PRIOR FILING DATE: 2001-02-09		
; PRIOR APPLICATION NUMBER: 60/269,308		
; PRIOR FILING DATE: 2001-02-16		
; Remaining Prior Application data removed - See File Wrapper or PALM.		
; NUMBER OF SEQ ID NOS: 78614		
; SOFTWARE: PatentIn version 3.1		
; SEQ ID NO 40681		
; LENGTH: 14067		
; TYPE: DNA		
; ORGANISM: Ureaplasma urealyticum		
US-10-282-122A-40681		
Query Match . 3.3%; Score 50.4; DB 13; Length 14067;		
Best Local Similarity 42.8%; Pred. No. 4;		
Matches 313; Conservative 0; Mismatches 416; Indels 3; Gaps 1;		
Qy	435 ATATGGATATTAATGTTTCTAGCACGTGTTAATAATGGAGCTGTATTAATTATGTTACATC 494	
Db	8784 AATGTTAGTTCTAATGATTCCTATAAATACTACAACAACAAGTAAATATGATATTGA 8843	
Qy	495 TGTAACTCCAAATGCTACTATTTCATGAAAAAGAAATACTGATCGACATGGGGAGATGGTGG 554	
Db	8844 TGGAAATCAAGAACTGAAATTAATAAAAAATTCAAATTCGTATATACTTCAATGACGG 8903	
Qy	555 TGGAAAACTGTAGATCAAAAAAGTACTCGTGGTGATACAGTCAAAATATACTATTATC 614	
Db	8904 TGAAGAAATTTAAGTGTATCAAAAAACATTCGCTTTGGGCTAATAAACCACTATAGTTTCGA 8963	
Qy	615 TTATAGAATGCAGTCAATTAATCATGTCAGAAAAAGTGTATCAATATGTTATAAGGA 674	
Db	8964 ATTGCTAAATTTAAACAAATATCGTAATAATACACTAAAGAAAGTAGAAGATTATTAATGA 9023	
Qy	675 TACTATGCCATCTGCTTCTGTAGTTGATTGTAACGAAGGGTCTTTATCAAGTAACTATTAC 734	
Db	9024 TGATATAAACATCAATAATTTTCCATTTAAAAATGGAATGACTGATTTGATTTATGT 9083	
Qy	735 TGATGGATCAGGGAATATTAACTCTPAACTCAAGGTCGCGAAAAAGCAACTCGGAAGTA 794	
Db	9084 TAATAAAACAATCGATTAGTATAAGTTCAATAATTGAGCCAATTCGAAGAGCGAAAAA 9143	
Qy	795 TAACCTGTTAGAGGAAAAATAATAATTTTCAGATTACTATTTCCGTGGGACGTACCAATAC 854	
Db	9144 TAATTTTACAATCAACACAAATTAGATTATTAATAATGATCCTGATTAATGTTTAAAGTAA 9203	
Qy	855 TCCAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTTTATAAGGGAATAAATAC 914	
Db	9204 TGAAGAAGAAGCAATCAATTAATTTATGTTGAGAAT---TTATCTGTAANAAGCAAAAGTTAA 9260	
Qy	915 AATCACAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACACAGGTTTCAGCTGATTT 974	
Db	9261 AATTGTAGGAGTCAAAAAATATTAGAAAACAACATTCATAATCTAGCATTTAAACCAAGA 9320	
Qy	975 ACCGAAAAATACAAACATTCGGACCATCAACCCCAATACCTAGCAATGATGACCCAGGTCA 1034	
Db	9321 TACTATTATTAAACAGTATTACTTTTATTAATAAACCAATAAAAAGCTCTACGAATATTGG 9380	
Qy	1035 AAAAGTAACAGTCAGGGATGCTCAAAATTAATCTATAAAAAAATTTGATGTTTCCCAAAAGC 1094	
Db	9381 AATAGATAATAGCAATAAAATATATACAATAACAATAACAATAAATAAATCCATTAATAAATTTGA 9440	
Qy	1095 TTCATTACAAGGTGCTATATTGTTGTTTTAAAGAAATGCTACGGGTCAATTTTCTTAAACTTTAA 1154	

Db	9441	TAATAATTTTAGTGTCTATGGTCTGTTCAAAGTGATACTAATAAACGTAATTTCTTGA	9500
Qy	1155	CGATACAAATAA	1166
Db	9501	ATATAATAATAA	9512

RESULT 11

US-10-087-464-42

Sequence 42, Application US/10087464

Publication No. US20030059436A1

GENERAL INFORMATION:

APPLICANT: Chishti, Athar

APPLICANT: Oh, Steven

APPLICANT: Liu, David

APPLICANT: Goel, Vikas

APPLICANT: Li, Xuerong

TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof

FILE REFERENCE: S1237/7019

CURRENT APPLICATION NUMBER: US/10/087,464

CURRENT FILING DATE: 2002-03-01

PRIOR APPLICATION NUMBER: US 06/272,930

PRIOR FILING DATE: 2001-03-02

NUMBER OF SEQ ID NOS: 59

SOFTWARE: PatentIn version 3.0

SEQ ID NO 42

LENGTH: 3996

TYPE: DNA

ORGANISM: Plasmodium falciparum

US-10-087-464-42

Query Match	3.3%	Score 50;	DB 15;	Length 3996;
Best Local Similarity	47.5%	Pred. No. 2.6;		
Matches 149;	Conservative 0;	Mismatches 165;	Indels 0;	Gaps 0;

Qy	366	TAATACTACACCAAGTTTCCACTGTTACTGAGTCACAAATAATGATCGTACTGAGGTTATTAA	425
Db	1881	TAATAATAATGATGAGAATTGTGATAATAATAAACCATATAATAATTAATAATAATAA	1940
Qy	426	TGTTCCCAATATGGATATTATTATGTTTCTAGCAGCTGTTTAATATGAGGCTGTAATTAT	485
Db	1941	TAATAATAATTATGTTGTAACAATAATAATAACAATAATAATAAGGCAATAATAATAA	2000
Qy	486	GGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAGAAATACTGATCGACATGGG	545
Db	2001	TGATGGAATGGTAGTAGTAATAATAATAATAATAATAATGATGATGACGAAAGAAAGAGA	2060
Qy	546	AGATGGTGGTGGAAAAAAGTGTAGATCAAAAACGCTACTCGGTTGGTGTATACAGTCAAAATA	605
Db	2061	TGATGAAGATGATAACAATAATAATAATAATGATGATAATAATGATGATAACGAAGAAAT	2120
Qy	606	TACTATTACTTTAAGAATCGAGTCAATTATTCATGGTACAGAAAAGTGTATCAATATGT	665
Db	2121	GGAAAGATAATGATGAAGATAACGATGAGTATAATAATAATGTAATGATAGTTATAAATATGA	2180
Qy	666	TATAAAGGATACTA	679
Db	2181	AGAAAAGATAGTA	2194

RESULT 12

US-10-206-576-219

Sequence 219, Application US/10206576

Publication No. US20030017495A1

GENERAL INFORMATION:

APPLICANT: Choi et al.

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 497

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

APPLICANT: Choi et al.  
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
 NUMBER OF SEQUENCES: 497  
 CORRESPONDENCE ADDRESS:  
     ADDRESSEE: Human Genome Sciences, Inc.  
     STREET: 9410 Key West Avenue  
     CITY: Rockville

**us-10-009-254-1.rnpb**

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35506
; LENGTH: 4997
; TYPE: DNA
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35506

Query Match      3.2%; Score 48.6; DB 13; Length 4997;
Best Local Similarity 43.5%; Pred. No. 5.8;
Matches 271; Conservative 0; Mismatches 349; Indels 3; Gaps 1;

QY   28 TTATGGCGAGTTTAGCACTTTCGATGCTGTATCACCAGTTACGCCGATAGCTTTTGCC 87
DB   1741 TCAGATGCAAAACAAGCAATTGATACAGCAGCTTTAAACCAAATAATCATTTAGTCAATAT 1800

QY   88 GCTGAGACAGGCACAATTACAGTTCGAAGATACTCAAAAAGGCGCAACCTATAAGACATAT 147
DB   1801 AACNATGAAGCCACTCAAGAAGAAAGATGTAGCTTTAGGCCAAAATTCATGAAGCAGCA 1860

QY   148 AAAGTTTTTGTGTCGAAATAGATTAATGCAATGTATCTGATTCGNATAAGATCGAGCT 207
DB   1861 AACAAGCTTAAAGCTGCCATAGATGCTGCAACTACAAATAATGCG--TGTAGATGAAGCA 1917

QY   208 TCTTTATTAATTCCTCAAGGTAAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAATTC 267
DB   1918 ACAACANATATACTACATTTATTTCTGAATACTTCTTGNTACGGTGAAGAAGCAGCA 1977

QY   268 CTTTTTCAGCAACTACTAATGAGGAGGAGAACATATGTAACATAAAAGATACCTGGTCA 327
DB   1978 GCAAGAAAAAGCGATTGATGATGCAGCAACTGCTGTAAAAAAAGACGCGATAAATAATACGTCA 2037

QY   328 GCANAATGAGATTCGACATGGCTTAATCTATATCAGTAACTACACCGATTTCCACT 387
DB   2038 GATGCAACAACAAGAAAGAATGACGCAANTAGCGAAAGTTGATGTCAGCTGTAAACAGCT 2097

QY   388 GTTACTCAGTCAAAATATGATGGTACTGAGGTTATTAATGTTTCCCATAATGATATTTAT 447
DB   2098 GCAAAACAAGCAATTACACAAGCAACAACAATAATGATAATGTAGACCAAGAACAAAAATAGC 2157

QY   448 TAGTTTTCTAGCACTGTTTAATAATGAGCTGTAATTTATGTTTACATCTGTAACTCCAAT 507
DB   2158 GGTACTTCGACTATTACTTGGTATACCAACGAGAAGTTACTTAAAAAACGACGACGACGAAA 2217

QY   508 GCTACTATTTCATGAAAAGAAATACTGTATGCCACATGGGGAGATGGTGGTGGAAAACTGTA 567
DB   2218 GCAATTCATGATGAGCTCGTTGCTTAAGAAAGCAGCAATTTGATACTGTTGCCGATGTACA 2277

QY   568 GATCAAAAAACGTACTCGTTGGTGTGATACAGTCAAAATATACTATTACTTATAGAAATGCA 627
DB   2278 GATGAAGAAAAACAGGCAGCTAAAGATAAAGTTGATGTGCTGAAGCTACCAAGCAAAAGCA 2337

QY   628 GTCAATTTATCATGTTACAGAAAA 650
```

```
Db      2338 GCGATTGATCAAGTACTACTACAAA 2360
      ||| ||| ||| ||| ||| ||| |||
RESULT 14
US-09-137-531-13
; Sequence 13, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
US-09-137-531-13
Query Match      3.2%; Score 48.4; DB 9; Length 3666;
Best Local Similarity 44.5%; Pred. No. 5.5;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY      302 ATGTAACTAAAAAGATACGCGTCAGCAATGAGATTGCGACATGGGCTAAATCTATAT 361
      ||| ||| ||| ||| ||| ||| |||
Db      2768 ATGTAAGTGAACAACACTGTTGATCTGCAACTGTTTCATTAAAGATAGTGCATAAAT 2827
      ||| ||| ||| ||| ||| ||| |||
QY      362 CAGCTAACTACACCACTGTTCCACTGTTACTGAGTCAAAATGAATGATGGTACTGAGGTTA 421
      ||| ||| ||| ||| ||| ||| |||
Db      2828 CATTATCTTTACATTAGTTGAAACTGGTGCTAATACAGGTGTATTTGCTACAACTGTT 2887
      ||| ||| ||| ||| ||| ||| |||
QY      422 TTAATGTTTCCCAATATGATATTTATGTTTCTAGCACTGTTAATAATGAGCTGTAA 481
      ||| ||| ||| ||| ||| ||| |||
Db      2888 AGCTGGTACATTAICTTTAACTGCTGGTACATTAAACAGTTACTTATGCAGATGCTA 2947
      ||| ||| ||| ||| ||| ||| |||
QY      482 TTATGTTTACATCTGTAATCCAAATGCTACTATTTTCATGAAAGAATACTGATGGACAT 541
      ||| ||| ||| ||| ||| ||| |||
Db      2948 AAAATGCTGAGGTGTTGCTGAAATATTTACTGTAGCTAACATTAAAGAAACTACTG 3007
      ||| ||| ||| ||| ||| ||| |||
QY      542 GGGAGATGGTGGGAAAACTGTAGATCAAAAAACGTACTCGTTGGTGTATACAGTCA 601
      ||| ||| ||| ||| ||| ||| |||
Db      3008 GAGCAATTACTTCTGATACATTTACACAAGGTGTATTACCATCAGCAGCTACAGCAGCTG 3067
      ||| ||| ||| ||| ||| ||| |||
QY      602 AATATACTATTACTTATAAGATGAGTCAATTTATCATGTTACAGAAAAAGTGTATCAAT 661
      ||| ||| ||| ||| ||| ||| |||
Db      3068 AATATACTTAAATCAATTGCTGAGATTATACATTTGCAACAGGTGAGGATTCATT 3127
      ||| ||| ||| ||| ||| ||| |||
QY      662 ATGTTATAAGGATACATATGCCATCTGCTTCTGTAGTTGATTTGAAACGAGGGTCTTATG 721
      ||| ||| ||| ||| ||| ||| |||
Db      3128 TAAATATTGATATGCTGGTCTCAAGTAAATTAACCTTAGCAGGTAAAAAAGGTGCACAAG 3187
      ||| ||| ||| ||| ||| ||| |||
QY      722 AAGTAACCTATTACT 735
      ||| ||| ||| ||| ||| |||
Db      3188 GTGTAGCTGATGCT 3201
      ||| ||| ||| ||| ||| |||

RESULT 16
US-09-137-531-7
; Sequence 7, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
```

```

; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-137-531-7

Query Match 3.2%; Score 48.4; DB 9; Length 4197;
Best Local Similarity 44.5%; Pred. No. 5.9;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAACATAAAAGATAGTGGTGCAGCAATGAGATGGGCTAAATCTATAT 361
DB 2952 ATGTAAGTGCAACAACTGTTGATAGTCAACTGTTTCATTAAAGATAGTGCAAAATAAT 3011

QY 362 CAGCTAAATACACAGATTTCCACTGTTACTGAGTCAAAATAATGATGGTACTCAGGTTA 421
DB 3012 CATTATCTTACATAGTTGAACTGGTCTAATACAGTTACTTATGCGATGCTA 3071

QY 422 TTAATGTTTCCAAATGATATATATATGTTTCTAGCACTGTTAAATATGAGCTGTAA 481
DB 3072 AAGCTGGTACATATCTCTTAACTGCTGTACATTAACAGTTACTTATGCGATGCTA 3131

QY 542 GGGGAGATGGTGGGAAAACTGTAGATCAAAAAAGTACTCGGTTGGTGATACAGTCA 601
DB 3192 GAGCAATTAATCTTGATACATTTACACAGGTGTATTAACATCAGCAGCTACAGCAGCTG 3251

QY 602 AATATACTATTACTTATAAGAAATGAGTCAATATCATGTTACAGAAAAGTGTATCAAT 661
DB 3252 AATATACTTCTAAATCAATTTGCTGAGATATACATTTGCAACAGGTGAAGGATTCATCT 3311

QY 662 ATGTTATAAGGATACATGCCATCTGCTGTAGTTGATTTGAACGAGGCTTTATG 721
DB 3312 TAAATATTGATATGCTGTGCTCAAGTAATTAATCTAGCAGGTAAAGAGTGCACAAG 3371

QY 722 AAGTAACTATTACT 735
DB 3372 GTGTAGCTGATGCT 3385

RESULT 17
US-09-137-531-8
; Sequence 8, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT: Expression of surface layer proteins
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-137-531-7

Query Match 3.2%; Score 48.4; DB 9; Length 4197;
Best Local Similarity 44.5%; Pred. No. 5.9;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAACATAAAAGATAGTGGTGCAGCAATGAGATGGGCTAAATCTATAT 361
DB 2952 ATGTAAGTGCAACAACTGTTGATAGTCAACTGTTTCATTAAAGATAGTGCAAAATAAT 3011

QY 362 CAGCTAAATACACAGATTTCCACTGTTACTGAGTCAAAATAATGATGGTACTCAGGTTA 421
DB 3012 CATTATCTTACATAGTTGAACTGGTCTAATACAGTTACTTATGCGATGCTA 3071

QY 422 TTAATGTTTCCAAATGATATATATATGTTTCTAGCACTGTTAAATATGAGCTGTAA 481
DB 3072 AAGCTGGTACATATCTCTTAACTGCTGTACATTAACAGTTACTTATGCGATGCTA 3131

QY 482 TTATGTTTCCAAATGATATATATATGTTTCTAGCACTGTTAAATATGAGCTGTAA 481
DB 3132 AAGCTGGTACATATCTCTTAACTGCTGTACATTAACAGTTACTTATGCGATGCTA 3131

QY 542 GGGGAGATGGTGGGAAAACTGTAGATCAAAAAAGTACTCGGTTGGTGATACAGTCA 601
DB 3192 GAGCAATTAATCTTGATACATTTACACAGGTGTATTAACATCAGCAGCTACAGCAGCTG 3251

QY 602 AATATACTATTACTTATAAGAAATGAGTCAATATCATGTTACAGAAAAGTGTATCAAT 661
DB 3252 AATATACTTCTAAATCAATTTGCTGAGATATACATTTGCAACAGGTGAAGGATTCATCT 3311

QY 662 ATGTTATAAGGATACATGCCATCTGCTGTAGTTGATTTGAACGAGGCTTTATG 721
DB 3312 TAAATATTGATATGCTGTGCTCAAGTAATTAATCTAGCAGGTAAAGAGTGCACAAG 3371

QY 722 AAGTAACTATTACT 735
DB 3372 GTGTAGCTGATGCT 3385

RESULT 18
US-10-373-667-3
; Sequence 3, Application US/10373667
; Publication No. US20030235841A1
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Koob, Michael
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 11000900101

```



```
; CURRENT APPLICATION NUMBER: US/10/373,667
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/181,585
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1037
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cDNA
; OTHER INFORMATION: comprising exons E, C, and A
US-10-373-667-3

Query Match          3.1%; Score 47.4; DB 16; Length 1037;
Best Local Similarity 47.2%; Pred. No. 4.8;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 218 TTCCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAAGTAAATTTAAATTTTACGA 277
DB 611 TCCTTCATGTTAGAAAACCTGGCTTTACTACTACTACTACTACTACTACTACTACTA 670

QY 278 CAACTACTAATGAGGGAGAAACATATGTAATAAAGAAAGATACCTGCGTCAGCAATGAGA 337
DB 671 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 730

QY 338 TTGCGACATGGCTAAATCTATATCAGCTAACTACAGCTAACTACACAGTTTCCACTGTTACTGAGT 397
DB 731 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 790

QY 398 CAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGATATTTTCTTA 457
DB 791 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850

QY 458 GCACGTGTTAATGAGAGCTGTAATATGTTTACATCTGTAATCTCAATGCTACTATTC 517
DB 851 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910

QY 518 ATGAA 522
DB 911 CTGCA 915

RESULT 19
US-10-373-667-1
; Sequence 1, Application US/10373667
; Publication No. US20030235841A1
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 11000900101
; CURRENT APPLICATION NUMBER: US/10/373,667
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/181,585
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-373-667-1

Query Match          3.1%; Score 47.4; DB 16; Length 1159;
Best Local Similarity 47.2%; Pred. No. 5.1;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 218 TTCCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAAGTAAATTTAAATTTTACGA 277
DB 423 TCCTTCATGTTAGAAAACCTGGCTTTACTACTACTACTACTACTACTACTACTACTACTA 482
```

```
QY 278 CAACTACTAATGAGGGAGAAACATATGTAATAAAGAAAGATACCTGCGTCAGCAATGAGA 337
DB 483 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542

QY 338 TTGCGACATGGCTAAATCTATATCAGCTAACTACAGCTAACTACACAGTTTCCACTGTTACTGAGT 397
DB 543 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602

QY 398 CAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGATATTTTCTTA 457
DB 603 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662

QY 458 GCACGTGTTAATGAGAGCTGTAATATGTTTACATCTGTAATCTCAATGCTACTATTC 517
DB 663 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722

QY 518 ATGAA 522
DB 723 CTGCA 727

RESULT 20
US-10-373-667-2
; Sequence 2, Application US/10373667
; Publication No. US20030235841A1
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 11000900101
; CURRENT APPLICATION NUMBER: US/10/373,667
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/181,585
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cDNA
; OTHER INFORMATION: comprising exons D, C, B, and A
US-10-373-667-2

Query Match          3.1%; Score 47.4; DB 16; Length 1471;
Best Local Similarity 47.2%; Pred. No. 5.7;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 218 TTCCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAAGTAAATTTAAATTTTACGA 277
DB 1045 TCCTTCATGTTAGAAAACCTGGCTTTACTACTACTACTACTACTACTACTACTACTACTA 1104

QY 278 CAACTACTAATGAGGGAGAAACATATGTAATAAAGAAAGATACCTGCGTCAGCAATGAGA 337
DB 1105 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1164

QY 338 TTGCGACATGGCTAAATCTATATCAGCTAACTACAGCTAACTACACAGTTTCCACTGTTACTGAGT 397
DB 1165 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1224

QY 398 CAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGATATTTTCTTA 457
DB 1225 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1284

QY 458 GCACGTGTTAATGAGAGCTGTAATATGTTTACATCTGTAATCTCAATGCTACTATTC 517
DB 1285 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1344

QY 518 ATGAA 522
DB 1345 CTGCA 1349
```

```

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9039
; LENGTH: 4050
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(4050)
; US-09-815-242-9039

Query Match      3.1%; Score 47.4; DB 9; Length 4050;
Best Local Similarity 46.1%; Pred. No. 9.5;
Matches 159; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 162 AGAATAGATATGCAATGATCTGATTCGAATTAAGATGGAGCTTCTTTAATTC 221
Db 333 AGAAAAAGAAAGTGATCAATCTACCACTGGAAATAAAGTTGAAGTTTCAACTGCCAAATC 392
QY 222 TCAAGGTAAGAAAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTTTTACCACAAAC 281
Db 393 AGATGAGCAAGCTTCACCAAAATCTACGAATGAAGATTTAAACACTAAACAACTATAAG 452
QY 282 TACTAATGGAGGGAGAACATATGTAACTAAAAAGATACCTGCGTCAGCAAAATGAGATTGC 341
Db 453 TAATCAAGAAAGCGTTTACAACTGATTTGCAAGAGAAATAAATCAGTGGTAAATTTCAACC 512
QY 342 GACATGGGCTAAATCTATATCAGTAACTACACAGCTTTCCACTGTTTACTGAGTCAA 401
Db 513 AACTAATGAGGAAACAAAGGTAGATGCCAAACTGAATCACTACATTAATGTTAA 572
QY 402 TAATGATGCTACTGAGGTTTATTAATGTTTCCCAATATGATATTTATGTTTCTAGCAC 461
Db 573 AAGTGATCTATCAAGAGTAAATGATGAACCTCTTGTGTGATAACAATAGTAATTCAAATAA 632
QY 462 TGTTAATAATGAGCTGTAATTTATGTTTACATCTGTAACTCCAAA 506
Db 633 TGAATAATATGCAGATATCAITTTTGCAAAAAGTACAGCACCTAA 677

RESULT 23
US-08-781-986A-454/c
; Sequence 454, Application US/08781986A
; Publication No. US20030054436A1

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4843
; LENGTH: 4047
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-815-242-4843

Query Match      3.1%; Score 47.4; DB 9; Length 4047;
Best Local Similarity 46.1%; Pred. No. 9.5;
Matches 159; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 162 AGAATAGATATGCAATGATCTGATTCGAATTAAGATGGAGCTTCTTTAATTC 221
Db 333 AGAAAAAGAAAGTGATCAATCTACCACTGGAAATAAAGTTGAAGTTTCAACTGCCAAATC 392
QY 222 TCAAGGTAAGAAAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTTTTACCACAAAC 281
Db 393 AGATGAGCAAGCTTCACCAAAATCTACGAATGAAGATTTAAACACTAAACAACTATAAG 452
QY 282 TACTAATGGAGGGAGAACATATGTAACTAAAAAGATACCTGCGTCAGCAAAATGAGATTGC 341
Db 453 TAATCAAGAAAGCGTTTACAACTGATTTGCAAGAGAAATAAATCAGTGGTAAATTTCAACC 512
QY 342 GACATGGGCTAAATCTATATCAGTAACTACACAGCTTTCCACTGTTTACTGAGTCAA 401
Db 513 AACTAATGAGGAAACAAAGGTAGATGCCAAACTGAATCACTACATTAATGTTAA 572
QY 402 TAATGATGCTACTGAGGTTTATTAATGTTTCCCAATATGATATTTATGTTTCTAGCAC 461
Db 573 AAGTGATCTATCAAGAGTAAATGATGAACCTCTTGTGTGATAACAATAGTAATTCAAATAA 632
QY 462 TGTTAATAATGAGCTGTAATTTATGTTTACATCTGTAACTCCAAA 506
Db 633 TGAATAATATGCAGATATCAITTTTGCAAAAAGTACAGCACCTAA 677

RESULT 23
US-09-815-242-9039
; Sequence 9039, Application US/09815242
```







```
QY 402 TAATGATGGTACTAGGTTTAATTAATGTTCCCAATATGGATATTAATTAATGTTTCTAGCAC 461
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
573 AAGTGATGCTATCAAGAGTAATGCTGAACACTCTGTTGATAACAATAGTAATTCAAATAA 632
QY 462 TGTTAATAATGAGCTGTGAATTAATGTTTACATCTGTAACTCCAAATGCTACTATTTCATGA 521
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
633 TGAATAATAATGAGATATCATTTTGGCAAAAAGTACAGCACCTTAAAGATTGTAATCAAG 692
QY 522 AAAGAATACTGATGCGACA 540
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 AATGCGTATGGCAGCAATA 711

RESULT 29
US-10-398-221-9/c
; Sequence 9, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 9
; LENGTH: 684707
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9

Query Match 3.0%; Score 45.8; DB 16; Length 684707;
Best Local Similarity 44.9%; Pred. No. 2.6e+02;
Matches 173; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 956 AACCAGGTTTCAGCTGATTTTACCAGAAAATACAAACATTTGCGACCATCAACCCCAATCTA 1015
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175405 AAGAGATTATCAATATAAAACGAAAGAGGAAATTTAAAAATGTTTAAAGTAGATCATA 175346
QY 1016 GCAATGATGACCCAGGTCAAAAGTAACAGTGAGGGATGGTCAAAATTACTATAAAAAAAA 1075
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175345 ATGATGTTTTACAAATGGAGTAGAAAATGGTACGTATGAGGTGGTTTTATACACGCCAA 175286
QY 1076 TTGATGGTTCCAAAAGCTTCATTACAGGTGCTATATTGTTTAAAGAAATGCTACGG 1135
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175285 ATGAAGATGCGACAAAACCGGAGCGGAGTTCAATTAATTTGATTTAATTTTCGTAATG 175226
QY 1136 GTCAATTTCTAAACTTTTAAACGATACAAATACCTTGAATGGGCGACAGAACTAATGCAA 1195
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175225 ATGTAATCAAAATTTCCGAATGCGCATATTTTCCCGAGTATGGAAGCAAAAGCAA 175166
QY 1196 CAGAATATACACAGGAGCAGATGGTATAATTAACATTAACAGCTTGAAGAGGTACAT 1255
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175165 CAAATGAATATAGTCAAAACCGCATTTAAATACATCGCTAAAGCAATCCAATTAACCTAACG 175106
QY 1256 ACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTTACATTTGTTAGATAACTCTCAGAAGG 1315
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175105 GCAAGATTAATATACATTTGGATGAATTTTAAAGACCTGTTAACTAAGACATGCGCAAG 175046
QY 1316 TTATTTTAGGAGATGGAGCCACTGA 1340
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175045 TTACTGTGAAAATGAAGAGTCTGA 175021
```

```
RESULT 30
US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
; US-10-398-221-2058

Query Match 3.0%; Score 45.8; DB 16; Length 3011208;
Best Local Similarity 44.9%; Pred. No. 4.6e+02;
Matches 173; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 956 RACCAGGTTTCAGCTGATTTTACCAGAAAATACAAACATTTGCGACCATCAACCCCAATCTA 1015
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2612988 AAGAGATTATCAATATAAAACGAAAGAGGAAATTTAAAAATGTTTAAAGTAGATCATA 2612929
QY 1016 GCAATGATGACCCAGGTCAAAAGTAACAGTGAGGGATGGTCAAAATTACTATAAAAAAAA 1075
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2612928 ATGATGTTTTCACAAATGGAGTAGAAAATGGTACGTATGAGGTGGTTTTATACACGCCAA 2612869
QY 1076 TTGATGGTTCCAAAAGCTTCATTACAGGTGCTATATTGTTTAAAGAAATGCTACGG 1135
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2612868 ATGAAGATGCGACAAAACCGGAGCGGAGTTCAATTAATTTGATTTAATTTTCGTAATG 2612809
QY 1136 GTCAATTTCTAAACTTTTAAACGATACAAATACCTTGAATGGGCGACAGAACTAATGCAA 1195
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2612808 ATGTAATCAAAATTTCCGAATGCGCATATTTTCCCGAGTATGGAAGCAAAAGCAA 2612749
QY 1196 CAGAATATACACAGGAGCAGATGGTATAATTAACATTAACAGCTTGAAGAGGTACAT 1255
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2612748 CAAATGAATATAGTCAAAACCGCATTTAAATACAAATCGCTAAAGCAATCCAATTAACG 2612689
QY 1256 ACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTTACATTTGTTAGATAACTCTCAGAAGG 1315
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2612688 GCAAGATTAATATACATTTGGATGAATTTTAAAGACCTGTTAACTAAGACATGCGCAAG 2612629
QY 1316 TTATTTTAGGAGATGGAGCCACTGA 1340
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2612628 TTACTGTGAAAATGAAGAGTCTGA 2612604

Search completed: July 27, 2004, 03:05:42
Job time : 739 secs
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 09:58:15 ; Search time 54 Seconds  
(without alignments)  
2626.645 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKQMIQSLVSLAFGMV.....GALLVIGAGIVLARRRLRS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2578	100.0	502	4	Aay72357 Virulent
2	2578	100.0	502	5	Aau77626 S. agalac
3	399	15.5	525	5	Abp65654 Bifidobac
4	334	13.0	627	2	Aay00118 Enterococ
5	334	13.0	627	5	Abp43337 E faecali
6	334	13.0	627	6	Abu88365 E. faecal
7	334	13.0	627	6	Abu13616 Enterococ
8	338	12.3	664	7	Adc97625 E. faeciu
9	279.5	10.8	705	5	Abp27284 Streptoco
10	274.5	10.6	560	2	AAY00119 Enterococ
11	274.5	10.6	560	5	Abp43338 E faecali
12	274.5	10.6	560	6	Abu88366 E. faecal
13	274.5	10.6	560	6	Abu13617 Enterococ
14	265.5	10.3	665	3	Abu00821 S. pneumo
15	265.5	10.3	665	3	AAY81627 Streptoco
16	258.5	10.0	614	5	Abb54071 Lactococc
17	244.5	9.5	554	5	Abp29802 Streptoco
18	244.5	9.5	581	5	Abp27373 Streptoco
19	234.5	9.1	341	7	Adc95757 E. faeciu
20	232	9.0	793	5	Abb47318 Listeria
21	232	9.0	793	6	Abu132765 Protein e
22	222.5	8.6	688	4	Aam16315 Peptide #
23	222.5	8.6	688	4	Abb35307 Peptide #
24	222.5	8.6	688	4	Aam28810 Peptide #
25	222.5	8.6	688	4	Abb30137 Peptide #

RESULT 1

AAY72357

ID AAY72357 standard; protein; 502 AA.

#### ALIGNMENTS

26	222.5	8.6	688	4	ABB20749	Abb20749	Protein #
27	222.5	8.6	688	4	AAM68511	Aam68511	Human bon
28	222.5	8.6	688	4	AAM56138	Aam56138	Human bra
29	222.5	8.6	688	4	ABG50180	Abg50180	Human liv
30	222.5	8.6	688	4	AAM04053	Aam04053	Peptide #
31	222.5	8.6	688	5	ABG38092	Abg38092	Human pep
32	220	8.5	560	5	AU77628	S. agalac	
33	220	8.5	563	4	AAY72358	Aay72358	Virulent
34	219	8.5	1222	3	AA01830	H. influe	
35	219	8.5	1228	3	AA01828	Haemophil	
36	219	8.5	1612	5	ABB47333	Listeria	
37	217	8.4	562	5	ABB47328	Listeria	
38	212	8.2	645	6	ASU45710	Protein e	
39	212	8.2	2551	6	ABU46159	Protein e	
40	208.5	8.1	660	7	ADC97088	E. faeciu	
41	205.5	8.0	1027	7	ADC97048	E. faeciu	
42	201.5	7.8	887	3	AAY81626	Streptoco	
43	200.5	7.8	893	6	ABU00820	S. pneumo	
44	200	7.8	1440	5	ABB54801	Lactococc	
45	199	7.7	508	7	ADC95704	E. faeciu	
46	199	7.7	1530	5	ABB47329	Listeria	
47	198.5	7.7	724	6	ABU17871	Protein e	
48	198	7.7	596	3	AAY99408	Human PRO	
49	198	7.7	596	4	AB66157	Protein o	
50	198	7.7	596	4	AAU29178	Human PRO	
51	198	7.7	596	4	AB87575	Human PRO	
52	198	7.7	596	5	ABG95900	Human sec	
53	198	7.7	596	6	ABU58554	Human PRO	
54	198	7.7	596	6	ABU88102	Novel hum	
55	198	7.7	596	6	ABU84417	Human sec	
56	198	7.7	596	6	ABR66291	Human sec	
57	198	7.7	596	6	ABR65681	Human sec	
58	198	7.7	596	6	ABU99621	Human sec	
59	198	7.7	596	6	ABU82860	Human PRO	
60	198	7.7	596	6	ABU89981	Novel hum	
61	198	7.7	596	6	ABR68230	Human sec	
62	198	7.7	596	6	ABU96283	Novel hum	
63	198	7.7	596	6	ABU92714	Human sec	
64	198	7.7	596	6	ABO08791	Human sec	
65	198	7.7	596	6	ABO02843	Human sec	
66	198	7.7	596	6	ABR74997	Human sec	
67	198	7.7	596	6	ABR94759	Human sec	
68	198	7.7	596	6	ABU85732	Human PRO	
69	198	7.7	596	6	ABU98892	Novel hum	
70	198	7.7	596	6	ABU98107	Novel hum	
71	198	7.7	596	6	ABU91813	Novel hum	
72	198	7.7	596	6	ABU89506	Human PRO	
73	198	7.7	596	6	ABU86347	Human sec	
74	198	7.7	596	6	ABU67560	Human sec	
75	198	7.7	596	6	ABU80588	Human PRO	
76	198	7.7	596	6	ABU90925	Novel hum	
77	198	7.7	596	6	ABO33984	Human sec	
78	198	7.7	596	6	ABR99506	Human sec	
79	198	7.7	596	6	ABR98896	Human sec	
80	198	7.7	596	6	ABO16419	Human sec	
81	198	7.7	596	6	ABR92319	Human sec	
82	198	7.7	596	6	ABO18960	Human sec	
83	198	7.7	596	6	ABR78381	Human sec	
84	198	7.7	596	6	ABU72001	Novel hum	
85	198	7.7	596	6	ABU85117	Novel hum	
86	198	7.7	596	6	ABO00256	Novel hum	
87	198	7.7	596	6	ABO11588	Human sec	
88	198	7.7	596	6	ABO02233	Human sec	
89	198	7.7	596	6	ABU88807	Novel hum	
90	198	7.7	596	6	ABU83502	Human sec	



XX AAY72357;  
 XX 24-APR-2001 (first entry)  
 XX Virulent group B Streptococcus agalactiae spbl protein.  
 DE Type III virulent group B; spbl; cell wall bound protein; antibacterial;  
 KW immunisation; group B streptococci; GBS infection; pneumonia; meningitis;  
 KW endocarditis; osteoarticular infection; bacteraemia; vaccine.  
 XX Streptococcus agalactiae.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Region 1..6  
 FT /note= "This region is hydrophilic and has six basic  
 FT amino acid residues"  
 FT Peptide 7..29  
 FT /label= Signal peptide  
 FT /note= "This region is hydrophobic and is rich in  
 FT proline"  
 FT Protein 30..502  
 FT /label= mature\_spbl\_protein  
 XX  
 XX MO200078787-A1.  
 XX  
 XX 28-DEC-2000.  
 XX  
 XX 21-JUN-2000; 2000WO-US017082.  
 XX  
 XX 21-JUN-1999; 99US-0140084P.  
 XX  
 XX (UTAH ) UNIV UTAH RES FOUND.  
 XX  
 XX Adderson E, Bohnsack J;  
 XX  
 XX WPI; 2001-102693/11.  
 DR N-PSDB; AAD02390.  
 XX  
 XX Polynucleotide from spbl and 2 genes derived from virulent Group B  
 PT streptococci, polypeptide encoded by the polynucleotide useful as vaccine  
 PT for immunizing a mammal against the streptococcal infection.  
 XX  
 XX Claim 56; Page 23-24; 34pp; English.  
 PS  
 XX The present sequence is spbl protein from type III virulent group B  
 CC Streptococcus agalactiae. The spbl protein has the characteristics of a  
 CC cell wall bound protein and has antibacterial activity. The N-terminus  
 CC of the spbl protein is a hydrophilic, basic stretch of 6 amino acids  
 CC followed by a 23 amino acid hydrophobic, proline rich core, consistent  
 CC with a signal peptide. The hydrophilic mature protein terminates in  
 CC atypical LPxTG domain that immediately precedes a hydro- phobic 20 amino  
 CC acid core and a short, basic hydrophilic terminus. The spbl protein are  
 CC used as a vaccine to immunise mammals against group B Streptococci (GBS)  
 CC infection (e.g. bacteraemia, pneumonia, meningitis, endocarditis and  
 CC osteoarticular infections). Determination of the gene products specific  
 CC to type III-3 GBS is useful for diagnosing mammals infected or colonised  
 CC by virulent GBS  
 XX  
 XX Sequence 502 AA,  
 SQ  
 Query Match 100.0%; Score 2578; DB 4; Length 502;  
 Best Local Similarity 100.0%; Pred. No. 1e-146;  
 Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKKKMIQSLVLASLAFGMVSPVTPAFATGCTITVQDTOKGATYKVKVFDPAEDNAN 60  
 DB 1 MKKKMIQSLVLASLAFGMVSPVTPAFATGCTITVQDTOKGATYKVKVFDPAEDIDNAN 60  
 QY 61 VSDSNKDGASYLIPQKGEAEYKASTDFNSLFTTTTNGGRYTVTKDTSASANEIATWAKSI 120  
 DB 61 VSDSNKDGASYLIPQKGEAEYKASTDFNSLFTTTTNGGRYTVTKDTSASANEIATWAKSI 120

QY 121 SANTTPSVSTVTESNNDGTEVINVSQYGYVYSSVTNNGAVIMVTSVTPNATHEKNVTDAT 180  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 121 SANTTPSVSTVTESNNDGTEVINVSQYGYVYSSVTNNGAVIMVTSVTPNATHEKNVTDAT 180  
 QY 181 WGDGGGKTVQDKTYSVGDVTKYTIYKNAVNYHGTEKYQYVVKDTMPSASVVDLNEGSY 240  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 181 WGDGGGKTVQDKTYSVGDVTKYTIYKNAVNYHGTEKYQYVVKDTMPSASVVDLNEGSY 240  
 QY 241 EYTIIDGSGNITITLQGGSEKATGKYNLLEENNFTITIPWAATNPTGNTQNGANDDDFFY 300  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 241 EYTIIDGSGNITITLQGGSEKATGKYNLLEENNFTITIPWAATNPTGNTQNGANDDDFFY 300  
 QY 301 KGINTITVTYTGVLKSGAKPGSADLPENTNIATINPNTSNDPPGQKVTVRDQIITIKKID 360  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 301 KGINTITVTYTGVLKSGAKPGSADLPENTNIATINPNTSNDPPGQKVTVRDQIITIKKID 360  
 QY 361 GSTKASLOGAIFVLKNATGQFLNFNDTNNVENGTEANATEYTTGADGIIITITGLKEGTY 420  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 361 GSTKASLOGAIFVLKNATGQFLNFNDTNNVENGTEANATEYTTGADGIIITITGLKEGTY 420  
 QY 421 LVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLVNPVTVENNKGTELPSTGGIGTITFY 480  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 421 LVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLVNPVTVENNKGTELPSTGGIGTITFY 480  
 QY 481 IIGAILVITGAGIVLVARRRLRS 502  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 481 IIGAILVITGAGIVLVARRRLRS 502  
 RESULT 2  
 AAU77626  
 ID AAU77626 standard; protein; 502 AA.  
 XX  
 AC AAU77626;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE S. agalactiae SbpI protein.  
 KW Extracellular matrix adhesion; Ema; group B streptococcus; GBS; vaccine;  
 KW SbpI; Sbp2; Rib; Lmb; C5a-ase; C protein alpha antigen;  
 KW neonatal bacterial infection.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200212294-A2.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 08-AUG-2001; 2001WO-US024795.  
 XX  
 PR 08-AUG-2000; 2000US-00634341.  
 XX  
 PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
 PA (UTAH ) UNIV UTAH RES FOUND.  
 XX  
 PI Adderson E, Bohnsack J;  
 XX  
 DR WPI; 2002-257465/30.  
 DR N-PSDB; ABK11585.  
 XX  
 XX New streptococcal matrix adhesion (Ema) polypeptides, useful as vaccines,  
 PT particularly for treating or preventing infections by virulent forms of  
 PT streptococci.  
 XX  
 PS Example 1; Page 150-152; 177pp; English.  
 PS  
 CC The invention relates to isolated streptococcal polypeptides, which  
 CC comprise the Streptococcus agalactiae extracellular matrix adhesion (Ema)  
 CC polypeptides EmaA, EmaB, EmaC, EmaD or EmaE, and their encoding nucleic  
 CC acids. Also included are the a vaccine comprising an Ema protein, an anti  
 CC -Ema antibody, a pharmaceutical composition comprising one or more Ema  
 CC proteins plus an anti Ema antibody, optionally in combination with at

CC least one antibody to a protein selected from Spb1 and Spb2, Rib, Lmb, C5a-ase or C protein alpha antigen, an immortal cell line producing an anti-Ema monoclonal antibody, a vector comprising the Ema nucleic acid, a host cell transformed with the vector, a nucleic acid vaccine comprising the vector and Ema homologous proteins and their encoding nucleic acids from additional bacterial species (*S. pneumoniae*, *S. pyogenes*, *Enterococcus faecalis* and *Corynebacterium diphtheriae*). The streptococcal polypeptides are useful as vaccines, particularly for treating or preventing infections by virulent forms of streptococci, especially group B streptococci (GBS) the most common cause of serious bacterial disease in neonates. The present sequence is the *S. agalactiae* Spb1 protein

XX Sequence 502 AA;

Query Match 100.0%; Score 2578; DB 5; Length 502;  
Best Local Similarity 100.0%; Pred. No. 1e-146;  
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKKKMIQSLVSLAFPGMAVSPVTPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNAN 60  
Db 1 MKKKMIQSLVSLAFPGMAVSPVTPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNAN 60  
Qy 61 VSDSNKDGASYLIPQKGAAYKASTDFNSLFTTTNGGRTYVTKKDTASANEIATWAKSI 120  
Db 61 VSDSNKDGASYLIPQKGAAYKASTDFNSLFTTTNGGRTYVTKKDTASANEIATWAKSI 120  
Qy 121 SANTTPVSTVESNNDGTEVINVSQYGYVSVSTVNGAVIMVTSVTPNATIHKNTDAT 180  
Db 121 SANTTPVSTVESNNDGTEVINVSQYGYVSVSTVNGAVIMVTSVTPNATIHKNTDAT 180  
Qy 181 WGDGGGKTVQKTSYSGVTYKTYIKYKAVNYHGTKEKYQYVVKDTMPSASVVDLNEGSI 240  
Db 181 WGDGGGKTVQKTSYSGVTYKTYIKYKAVNYHGTKEKYQYVVKDTMPSASVVDLNEGSI 240  
Qy 241 EVTTIDGSGNITTLTQSGEKATGKYNLLENNFTTIPWAATNPTGNTONGANDDFY 300  
Db 241 EVTTIDGSGNITTLTQSGEKATGKYNLLENNFTTIPWAATNPTGNTONGANDDFY 300  
Qy 301 KGINTITVYTVGLKSGAKPGSADLPENTNIATINPNTSDPGQKTVVRDQITIKKID 360  
Db 301 KGINTITVYTVGLKSGAKPGSADLPENTNIATINPNTSDPGQKTVVRDQITIKKID 360  
Qy 361 GSTKASLOGAIFVLKNATGQFLNFDNTNNVENGTEANATEYTTGADGIITITGLKEGTYI 420  
Db 361 GSTKASLOGAIFVLKNATGQFLNFDNTNNVENGTEANATEYTTGADGIITITGLKEGTYI 420  
Qy 421 LVEKKAPGLYNLLDNSQKVLGDGATDTTNSDNLVNPVENNKTELPSTGGIGTTIFY 480  
Db 421 LVEKKAPGLYNLLDNSQKVLGDGATDTTNSDNLVNPVENNKTELPSTGGIGTTIFY 480  
Qy 481 IIGAILVIGAGIVLVARRLRS 502  
Db 481 IIGAILVIGAGIVLVARRLRS 502

RESULT 3  
ID ABP65654 standard; protein; 525 AA.  
XX AC ABP65654;  
XX AC ABP65654;  
DT 19-NOV-2002 (first entry)  
XX Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:398.  
DE DE Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KW antidiarrheic; antibacterial; inhibitor of salmonella; detection;  
KW identification; lactic acid bacteria; diarrhoea; pathogenic bacteria;  
KW rotavirus; food composition; pharmaceutical composition.  
XX Bifidobacterium longum.  
OS Bifidobacterium longum.  
XX EP1227152-A1.

XX 31-JUL-2002.  
PD 30-JAN-2001; 2001EP-00102050.  
PF 30-JAN-2001; 2001EP-00102050.  
XX 30-JAN-2001; 2001EP-00102050.  
PR (NEST ) SOC PROD NESTLE SA.  
PA WPI; 2002-668397/72.  
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as a probe or primer for detecting and/or identifying Bifidobacterium longum in a biological sample.  
XX Claim 3; SEQ ID NO 398; 80pp; English.

The present invention describes a polynucleotide (I) comprising a sequence of a Bifidobacterium genome selected from the nucleotide sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at least 90% identity or which hybridises with the sequences given in ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a fusion protein, comprising a sequence selected from 1097 sequences given in ABQ81842 to ABQ81843 ligated in frame to a polynucleotide encoding a heterologous polypeptide. (I) has antidiarrheic and antibacterial activities, and can be used as an inhibitor of Salmonella. (I) (which is a probe) is useful for the detection and/or identification of Bifidobacterium longum in a biological sample. A carrier containing the lactic acid bacterium Bifidobacterium longum NCC2705 (NCM 1-2618) can be used for preventing and/or treating diarrhoea brought about by pathogenic bacteria and/or rotavirus. The carrier is a food composition selected from milk, yogurt, curd, cheese, fermented milks, milk based fermented products, ice-creams, fermented cereal based products, milk based powders, infant formula, pet food or a pharmaceutical composition selected from tablets, liquid bacterial suspensions, dried oral supplement, wet oral supplement, dry tube feeding or wet tube feeding. (I) is useful in DNA arrays or chips to carry out analysis of the expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent Bifidobacterium related nucleotide sequences given in the Sequence Listing from the present invention but not mentioned further within the specification. N.B. The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office

Sequence 525 AA;

Query Match 15.5%; Score 399; DB 5; Length 525;  
Best Local Similarity 27.3%; Pred. No. 7.3e-16;  
Matches 160; Conservative 72; Mismatches 198; Indels 156; Gaps 26;

Qy 3 KKMISQSLVSLAFPGMAVSPVTPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNA 59  
Db 6 RKVAAGVLAATMLGIAGLGTASAEATGTLVTSDDAFNGKYNAYQMFSPDAA 65  
Qy 60 NVSDSNKDGASYLIPQKGAAYKASTDFNSLFTTTNGGRTYVTKKDTAS 109  
Db 66 -----GQATYTLASPDWPNFKNADLGTITVTGDANVSEKAYDYVYSL 109  
Qy 110 -----ANEIATWAKSISANTTPVSTVTSNNDGTEVINVSQY-----GYYY 150  
Db 110 NQDMKLAEFATKASNAKAGNNVSAVNPAAATKASG-----SQYVATFANPLGYYV 163  
Qy 151 VS-----STVN-----NGAVIMVTSVTPNATIH-----EKNTDA-TWGDG-----GGKTVDDQK 192  
Db 164 VSPAGSTSNKRHTDAMLVNVTATTKNLKSEYPTVDKTDADKGDGSAIGSKVNFQL 223  
Qy 193 TYSVGDVTVKYITTYKNAVNYHGTKEKYQYVVKDTMPSASVVDLNEGSEVITITDGSNIT 252  
Db 224 KSKVPDTSVT-----NYVFKIVDTLSAG--LDFFNN--DVTVKVGDATLT 264  
Qy 253 TLTGSEKATGKYNLLENNFTTIPWAATNPTGNTONGANDDFYKGINITVTVTG 312  
Db 265 ATTDYSVTTRGK-----TVTI-----DLSNVYKTDNASKAGKGLVTVSA 304

Qy	313	VLKSGAKPGSADL--PENTNIATI--NPNTSNDPDGQKVTVRGQGIT-----IKI---DGS	362
Db	305	TLNENAFVGTDPQNNPGLNSAKVQYSGNPSEENIGESTPSETHSYTFNFKLKIYKEGD	364
Qy	363	TRASLOGAIFVLKNATGQFLNF-----NDTNVNMGTGANATAYTTTGADGII	409
Db	365	TENALAGAFQLLDSDKTVISLVKSDNVYRPAKTSOTDEV-----TEVETPATGII	416
Qy	410	TITGLKEGYIYVEKAPLPGYNLLDNSQKVLGDCATDTTNS--DNLLVN---ET-----	459
Db	417	EFTGLKAGYIYLKETFPAPKGYNKLSDPVKVTINATINKTIGALBSWTVNGSAPTADVTP	476
Qy	460	---VENNKTELPSTGGIGTTTFYIIGAILV-IGAGIVLVARRRKL	501
Db	477	VVKIENKKGALLPDTGGMGVLTFTVFGLVIALGAGWYVKSNRKSR	522

## RESULT 4

RESULT 4  
AAY00118  
ID AAY00118 standard; protein: 627 AA.

Query Match 13.0%; Score 334; DB 2; Length 627;  
Best Local Similarity 26.3%; Pred. No. 7.3e-12;  
Matches 150: Conservative 79; Mismatches 223; Indels 1

[illegible]

## RESULT 5

RESULT 5  
ABP43337  
ID ABP43337 standard; protein: 627 AA.

New genes and polypeptides from *Enterococcus faecalis*, useful as vaccines for preventing, treating or attenuating an infection caused by a member of the *Enterococcus* genus in an animal, particularly *E. faecalis*.

PS	Claim 9; Page 93; 255pp; English.	
XX	The present invention provides the protein and coding sequences of a	
CC	number of polypeptides from Enterococcus faecalis. The proteins can be	
CC	used as vaccines for preventing or attenuating an infection caused by a	
CC	member of the Enterococcus genus in an animal, particularly E. faecalis.	
CC	The polynucleotide is also useful for preventing or treating E. faecalis	
CC	infection. The present sequence is a protein of the invention	
XX		
SQ	Sequence 627 AA;	
	Query Match 13.0%; Score 334; DB 5; Length 627;	
	Best Local Similarity 26.3%; Pred. No. 7.3e-12;	
	Matches 150; Conservative 79; Mismatches 223; Indels 118; Gaps 26;	
QY	3 KMIQSLVASLAFGMVSPVTPPIAFAA--ETGTTIVQDTQKGATYKAYKVFDAIDNAN 60	
Db	100 KQAVQSL-----TPGKPVAGTTDANGNVTVQLPKKQNGKDAVYTIKEPKEGV 148	
QY	61 VSDSNKGA--SYLIPQKGEAYKASTDFNSLF-----TTTTNGGRTVYTKDTSANE- 112	
Db	149 VAATNMVAFVYEMIKQDGSYKYGTBELAVVHIYPKNVAVNDGSLHVKVGTAE-NEG 207	
QY	113 -----IATWA-----KSIANTTPV--STVTESNND 136	
Db	208 LNGAEFVISKSESGPGTVKYIQGVKDGLYTWTDKQAKRFTIGKSYEIGENDFTEAEN- 266	
QY	137 GTEVINVS--QGYYYVVS--STVNVGAVIMVTSVTP-----NATIHEK--NTDATWGDG 184	
Db	267 GTGELTVKNLVGSLYILEBVKAPNAELIENQTKPTFTIANNQTPVEKTVKNDISKVDK 326	
QY	185 GGTVDQKTYSGDVTYKTYITKYNVNVHGT-----KVYQYVYKDTMPSASVVD-LNEGS 239	
Db	327 TTPSLDGDVAIGEKIKYISVNIPLGIADKEGDANKYVKFNLVDKHAALTFDNTVSGE 386	
QY	240 YEVTTIDSGNITLTQSEKATGKYNLLENNNTIIT--PWAATNPTGNTQNGANDD 297	
Db	387 YAYALYDGDVTL-----APENYQVTEQANGFTAVNPAYIPTLTPGG----- 428	
QY	298 PFYKGINTITVYTGVLKSGAKPGSADLPENTNIATINPNTSDPDGQKTVRDRGQITIK 357	
Db	429 -----TLKFVYFPHLNEKADPTKG-----FKNEAVNDNGHTDDQTPPTVEVVTGKRFI 477	
QY	358 KIDGSTKA--SLQGAIFVLKNA---TGQFLNFNDTNN-VGW-GTEANATEYTTGADGIIT 410	
Db	478 KVDGVTATQALAGASFVVRDQNSDTANYLKIDETTKAATWVKRAEATFTTTADGLVD 537	
QY	411 ITGLKECTYLYEKKAPLGYNLLDNSQKVLGDGATDITNSDNLVNPVTEN-NKGTELP 469	
Db	538 ITGLXGYTYLEETVAPDDYVLLTNRIEFVNVNEQSYGTT--ENLVSPEKVPNKHKGTL-P 594	
QY	470 STGGIGTTIFYIIGAILVIGAGILVARRR 499	
Db	595 STGGKGIYVILGSGAVLLIAGVYFARRRK 624	
RESULT 6		
ABU88365		
ID	ABU88365 standard; protein; 627 AA.	
XX		
AC	ABU88365;	
XX		
DT	07-JUL-2003 (first entry)	
XX		
DE	E. faecalis novel protein #109.	
XX		
KW	Vaccine; endocarditis; bacteraemia; urinary tract infection; UTI;	
XX	intraabdominal infection; soft tissue infection; neonatal sepsis.	
OS	Enterococcus faecalis.	
XX		
PN	US2003017495-A1.	
XX		
PD	23-JAN-2003.	
XX		
PF	29-JUL-2002; 2002US-00206576.	
XX		
PR	06-MAY-1997; 97US-0044031P.	
PR	16-MAY-1997; 97US-0046655P.	
PR	14-NOV-1997; 97US-0066009P.	
PR	04-MAY-1998; 98US-0007103S.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Choi GH, Bailey C, Hromockyj A, Kunsch CA;	
XX		
WPI	2003-416890/39.	
DR	N-PSDB; ACA88057.	
XX		
PT	New nucleic acid molecules and polypeptides from Enterococcus faecalis,	
PT	useful as vaccines for preventing or attenuating an enterococcal	
PT	infection in an animal, or for identifying Enterococcus faecalis in	
PT	biological samples.	
XX		
PS	Claim 12; Page; 40pp; English.	
XX		
CC	The invention relates to a new isolated nucleic acid molecule comprising	
CC	a polynucleotide isolated from Enterococcus faecalis appearing as	
CC	ACA87949-ACA88196 (or sequences complementary to them or 95% identical to	
CC	them). Also included are the proteins encoded by the above nucleic acids,	
CC	making a recombinant vector (comprising inserting the isolated nucleic	
CC	acid molecule cited above into a vector), a host cell comprising the	
CC	vector (used to produce the protein), an isolated antibody specific for	
CC	the polypeptides, a hybridoma that produces the antibody, an isolated	
CC	polypeptide antigen comprising an amino acid sequence of an Enterococcus	
CC	faecalis epitope listed in the specification, a vaccine comprising one or	
CC	more E. faecalis polypeptides (and a pharmaceutical diluent, carrier or	
CC	excipient) where the polypeptide elicits protective antibodies in an	
CC	animal to a member of the genus Enterococcus; preventing or attenuating	
CC	an infection caused by a member of the genus Enterococcus in an animal	
CC	comprising administering to the animal the polypeptide and detecting	
CC	Enterococcus nucleic acids in a biological sample. The E. faecalis	
CC	nucleic acid molecules and polypeptides are useful as vaccines for	
CC	preventing or attenuating an enterococcal infection in an animal (e.g.	
CC	endocarditis, bacteraemia, urinary tract infection (UTI), intraabdominal	
CC	infection, soft tissue infection and neonatal sepsis). The polypeptides	
CC	are also useful for detecting Enterococcus aureus in immunoassays, as	
CC	epitope tags, as molecular weight markers, or for generating antibodies	
CC	that specifically bind E. faecalis polypeptides. The nucleic acid	
CC	molecules are also useful as probes for gene mapping, or for identifying	
CC	E. faecalis in biological samples. The kit and methods are useful for	
CC	detecting Enterococcus antibodies or nucleic acid molecules in a	
CC	biological sample. The present sequence is a novel E. faecalis	
CC	polypeptide of the invention. Note: The sequence data for this patent did	
CC	not form part of the printed specification, but was obtained in	
CC	electronic format directly from the USPTO at	
CC	seqdata.uspto.gov/sequence.html?docID=20030017495	
XX		
SQ	Sequence 627 AA;	
	Query Match 13.0%; Score 334; DB 6; Length 627;	
	Best Local Similarity 26.3%; Pred. No. 7.3e-12;	
	Matches 150; Conservative 79; Mismatches 223; Indels 118; Gaps 26;	
QY	3 KMIQSLVASLAFGMVSPVTPPIAFAA--ETGTTIVQDTQKGATYKAYKVFDAIDNAN 60	
Db	100 KQAVQSL-----TPGKPVAGTTDANGNVTVQLPKKQNGKDAVYTIKEPKEGV 148	
QY	61 VSDSNKGA--SYLIPQKGEAYKASTDFNSLF-----TTTTNGGRTVYTKDTSANE- 112	
Db	149 VAATNMVAFVYEMIKQDGSYKYGTBELAVVHIYPKNVAVNDGSLHVKVGTAE-NEG 207	
QY	113 -----IATWA-----KSIANTTPV--STVTESNND 136	
Db	208 LNGAEFVISKSESGPGTVKYIQGVKDGLYTWTDKQAKRFTIGKSYEIGENDFTEAEN- 266	













XX AC ABU00821;  
 XX DT 23-OCT-2003 (revised)  
 XX DT 11-FEB-2003 (first entry)  
 XX DE S. pneumoniae type 4 strain protein from coding region #389.  
 XX DE  
 XX DE Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
 KW gene therapy; vaccine.  
 XX OS Streptococcus pneumoniae; type 4 strain.  
 XX PN WO200277021-A2.  
 XX PD 03-OCT-2002.  
 XX PF 27-MAR-2002; 2002WO-1B002163.  
 XX PR 27-MAR-2001; 2001GB-00007659.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Massignani V, Tettelin H, Fraser C;  
 XX DR WPI; 2003-040579/03.  
 XX DR N-PSDB; ABX06101.  
 XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX PS Claim 1; SEQ ID NO 778; 56pp; English.  
 XX CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target  
 CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX SQ Sequence 665 AA;

Query Match 10.3%; Score 265.5; DB 6; Length 665;  
 Best Local Similarity 24.4%; Pred. No. 1e-07;  
 Matches 164; Conservative 63; Mismatches 227; Indels 219; Gaps 33;

QY 9 LLVASLAFGMAVSPVPIAFAAETGTITVQDTQKGAATYKAYKVFDAEIDNANVSDSNKDG 68  
 |||||  
 Db 15 LITASSLSFSAAT-----VFAAGTTTTVTVHKLLATDGDMDKIANELETGNYA--GNKVG 67  
 |||||  
 QY 69 ASYLIPQGEAEYKASTDFNSLFTTTN-----GGRTVTVTKKDTASANEIATWAKSISAN 123  
 |||||  
 Db 68 ---VLP-----ANAKEIAGVMFVNTNNEIIDENGQTLGVNIDP-----QTFKLSGAMP 114  
 |||||  
 QY 124 TTPVSTVTESNNDGTEVINV---SOVGYVYVSS---TVNNGAVIMVTSVTP----- 168  
 |||||  
 Db 115 ATAMKLUAEAGAKENTANLPAKYKIYEIHSLSYVGEDGATLTGSKAVFIEIPLND 174  
 |||||  
 QY 169 --NATIEHKNTDA--TWGDDGGKT-----VDQKT-----YSVGDTVKYTIITYK--NAVNY 212  
 |||||  
 Db 175 VVDAHVYPKNTAKPKIDKDFKXANPDTPRVDKDTPVNHQVGDVVEIYVTKIPALANY 234  
 |||||  
 QY 213 -----HGETKYV-----QVVIKDT-----MPSASVVDLNEG-- 238  
 |||||  
 Db 235 ATANWSDRMTEGLAFNKGTVKVTVDVALEAGDYALTAVATGFDLKLDTAGLAKVNDQNA 294  
 |||||  
 QY 239 -----SYEVTITDGS-----GNITLTQSEKATG---KYNLLENNNTTITPWA-A 282  
 |||||  
 Db 295 EKTVKITYSATLNDKAIVEVPESNDVTFYGNPDHGNTPKPNKPNENGDLTLTKTWDA 354  
 |||||  
 QY 283 TNTPTG-----NTQNGANDDFYKGINTI-----TVTYTGVLKS-----G 317  
 |||||  
 Db 355 TGAPIPAGAEATFDLVNAQTG-----KVQTVTLTDTKNTVTNGLDKNTEYKFVRS 407  
 |||||  
 QY 318 AKPSADLPENTNIATI-----NPNTSNDPQKQVTVRDQGIITIKIDGSTKASLQGA 371  
 |||||  
 Db 408 IKGSADYQEIITTAAGEIAVKMKDENPKPLDPTBPVVYTGKFKVNDKDNR--LAGAE 465  
 |||||  
 QY 372 FVLKNA--TGQFL-----  
 |||||  
 Db 466 FVIANADWAGVYLARKADKVSQBEKQLVVYTKDALDRAVAAYNALTAAQQTQOEKVKDK 525  
 |||||  
 QY 383 ---NFN-----DTNNVEMGTE---ANATEYTTGADGIITITGLKEGTYYLVEKKAPLGYN 431  
 |||||  
 Db 526 AQAAYNAAVIAANNAFVADKDNENVVKLVSDAQGRFEITGLLAGTYVLEETKQAGYA 585  
 |||||  
 QY 432 LLDNSQV-----ILGDGATDTTNSDNLVNPTVNNKGTBLPSTGGIGTTFIYIG 483  
 |||||  
 Db 586 LLSRQKFEVATSYATSGQIEYTAGSGK--DDATKVNKKITIPQTGGIGTTFIYAVAG 643  
 |||||  
 QY 484 AILVIGAGIVLVA 496  
 |||||  
 Db 644 AAIM---GIAVYA 653

## RESULT 15

AA81627  
 ID AA81627 standard; protein; 666 AA.

XX AC AA81627;

XX DT 24-MAY-2000 (first entry)

XX DE Streptococcus pneumoniae type 4 protein sequence #127.

XX KW Streptococcus pneumoniae; vaccine; screening; protein antigen;  
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
 KW pneumococcal disease.

XX OS Streptococcus pneumoniae.

XX PN WO200006737-A2.

XX PD 10-FEB-2000.

XX PF 27-JUL-1999; 99WO-GB002451.

XX PR 27-JUL-1998; 98GB-00016337.

PR	19-MAR-1999;	99US-0125164P.	
XX	(MICR-) MICROBIAL TECHNIKS LTD.		
XX	Gilbert CFG, Hansbro PM;		
XX	WPI; 2000-195300/17.		
XX	New Streptococcal protein, useful as a vaccine, for diagnosis of		
PT	PT pneumococcal diseases and for screening agents capable of antagonizing or		
PT	inhibiting expression of the protein.		
XX	Claim 1; Page 85; 108pp; English.		
XX	AA051501 to AA051679 represent specifically claimed protein sequences		
CC	isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent		
CC	specifically claimed nucleotide sequences isolated from S. pneumoniae.		
CC	The sequences have antibacterial and antiinflammatory properties. The		
CC	protein sequences, and fragments of them, are useful as immunogens and/or		
CC	antigens. The nucleotide sequences can be used in vaccines and in		
CC	diagnostic assays. The proteins and nucleotides can be useful for the		
CC	detection and diagnosis of S. pneumoniae. The protein sequences are also		
CC	useful for screening an agent capable of antagonising, inhibiting or		
CC	interfering with the function or expression of the proteins in which the		
CC	agent is useful for treatment or prophylaxis of S. pneumoniae infection		
CC	and meningitis. AA05591 to AA05614 represent primers used in the		
CC	exemplification of the present invention		
XX	Sequence 666 AA;		
XX	Query Match	10.3%; Score 265.5; DB 3; Length 666;	
XX	Best Local Similarity	24.4%; Pred. No. 1e-07;	
XX	Matches 164; Conservative	63; Mismatches 227; Indels 219; Gaps 33;	
QY	9	LLVASLAFGNVSPVPIAFATGTTVDQTKGATYKAYKVFDAEIDNANVSDNKG	68
DB	15	LLTASSLFSAA-----VFAAGTTTSVTHKLLATDGMCKIANELETGNYA-GNKVG	67
QY	69	ASYLIPQKGEAEYKASTDFNSLFTTTN-----GGRYTVTKKDTASANEIATWAKSISAN	123
DB	68	---VLP-----ANAKEIAGVMFVNTNNEIIDENGQTLGNIDP-----QIFKLSGMP	114
QY	124	TPVSTVTSNNDGTEVINV--SQYGYVYSS-----TVNGAVIMVTSVTP-----	168
DB	115	ATAMKLLTEAGAKFNKTANLPAKYKIYETHSLSTYVGEDGATLTGSKAVPIEIELPLND	174
QY	169	--NATHEKNTDA--TWGPGGCKT-----VDQKT--YSGVDTVKYITTYK--NAVNY	212
DB	175	VDAHVYPRKTEAKPKIDKDFGKANPDPTRVDRKDTFVNHQGVGVVEYELVTKIPALANY	234
QY	213	-----HGTEKVV-----QYVINDT-----MPSASVVDLNEG--	238
DB	235	ATANSDRMTEGLAFNKGTVKTVVDVALEAGDYALTEVATGFDLKLTDAGLAKVNDQNA	294
QY	239	-----SYEVTITDGS-----GNITTLTQSEKATG-----KYNLBNENNFTITIPWA-A	282
DB	295	EKTVKITYSATLNDKAIVEPESNDVTFNYGNPNPDHGNTPKPNKPNENGDLTLTKTWVDA	354
QY	283	TNTPGT-----NTQNGANDFPYKGINTI-----TVTYTGVLKS-----G	317
DB	355	TGAPIPAGAEATFDLNAQTG-----KVQVITLTDKNVITVNLGNLKNTEYKVFERS	407
QY	318	AKPGSADLPENTNIAT-----NPNTSNDPQKQVTRDQGITIKIKIDGSTKASIQGAI	371
DB	408	IKGYSADYQBITTAGETIAVKNWKENPKPLDPEPKVTVYGGKFKVKNVDKDNR--LAGAE	465
QY	372	FVLKNA--TCQFL-----	382
DB	466	FVTANADNAGQYLARKADKQVSEBKQLVWTTKDALDRAVAYNALTAQQOQOQEKVKDK	525
QY	383	---NFN-----DTNNVEWGPTE---ANATEVTTGADGIITITGLKEGTYYLVEKKAPLGN	431
DB	526	AQAYNAAVTAANNAFEWADKDNENVVVLVSDAQGFETGLLAGTYYLEETKQAGYA	585
QY	432	LLDNSQKV-----ILGDGATDTTNSDNLVLNFTVENNKGTETLPSTGGTGTTFYIIG	483
DB	586	LLTSRQKEFVATSYSATGQIEYTAGSGK--DDATKVVNKKITIPQTGGTGTTFYIIFAVAG	643
QY	484	AILVIGAGILVA 496	
DB	644	AAIM---GIAVYA 653	
XX	RESULT 16		
XX	ABB54071		
XX	ID	ABB54071 standard; protein; 614 AA.	
XX	AC	ABB54071;	
XX	DT	29-AUG-2003 (revised)	
XX	DT	16-MAY-2002 (first entry)	
XX	DE	Lactococcus lactis protein yhgE.	
XX	KW	Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.	
XX	OS	Lactococcus lactis; IL1403.	
XX	PN	FR2807446-A1.	
XX	PD	12-OCT-2001.	
XX	PF	11-APR-2000; 2000FR-00004630.	
XX	PR	11-APR-2000; 2000FR-00004630.	
XX	PA	(INRG ) INRA INST NAT RECH AGRONOMIQUE.	
XX	PI	Bolotine A, Sorokine A, Renault P, Ehrlich SD;	
XX	DR	WPI; 2002-043418/06.	
XX	PT	New nucleotide sequence useful in the identification of Lactococcus	
XX	PT	lactis and related species.	
XX	PS	Claim 6; SEQ ID NO 773; 2504pp; French.	
XX	CC	The present invention is related to a Lactococcus lactis nucleotide	
XX	CC	sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucle	
XX	CC	acid sequence is useful in the detection and/or amplification of nucle	
XX	CC	acid sequence, particularly to identify Lactococcus lactis or related	
XX	CC	species. The proteins of the invention are useful for the biosynthesis or	
XX	CC	biodegradation of a composition of interest. The invention helps research	
XX	CC	in lactic bacteria, particularly useful in the production of yogurt and	
XX	CC	cheese. Note: The sequence data for this patent is based on equivalent	
XX	CC	patent WO200177334 (published 18-OCT-2001) which is available in	
XX	CC	electronic format directly from WIPO at	
XX	CC	ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to	
XX	CC	standardise OS field)	
XX	SQ	Sequence 614 AA;	
XX	Query Match	10.0%; Score 258.5; DB 5; Length 614;	
XX	Best Local Similarity	24.5%; Pred. No. 2.4e-07;	
XX	Matches 145; Conservative	57; Mismatches 184; Indels 207; Gaps 31;	
QY	51	VFDAEIDN-----ANVSDSNKDGASYLIPOGKEAEYKASTDFNSLFTTTNGGRTYVT	103
DB	75	VNDSGIANNQTCGAGNTGSTRKNGSQTTLEGSTPSTMANVTFSATKYVGT-GVPTGVT	133
QY	104	K---KDTASANEIATWAKSISANT-----TPVSTVTSNNDGTEVINVSQ-----	145
DB	134	DPSFTDTTTSVAPVTTDASGLADFTGLTGYYLHFQVTTVNGITTVGDFIVQVSHEDSOAG	193
QY	146	---XGYVVSSTVNVNNAVIMVTSVT-----PNATHE-----KNITD--	178

```
Db 194 IVNVPKLDMSSAGLG-----TSATTNADNHNFGTNPQIANPNATGNSDOTLNTDNN 248
Qy 179 -----ATWGGGGKTVQDKYSGVDFVKYTIYKNAV-----NYHGTEKYV-QXV 222
Db 249 AGNENLANGTWTNGSDNQ-NTTTAAAGTNVNNV---NTVPDSSQTNNNGTGTGTGYI 304
Qy 223 IKDTHP-----SASVVDLNEGS-----YEVTTDSCGNIT--TTLTQG 257
Db 305 VTDQLPNNLVNLSVTSTVIIVNTNGSGTKVGTLTPTTPTDYTIT-NDGNGKIVVTLTTAG 363
Qy 258 SEKATGKYNLLBNENNFTIIPWAATNPTGNTONGANDFFYKGINITITVYTGVLKSG 317
Db 364 QHRAASL--LGSADGALNLIIP-STVKSAGSATSAT-----TITNAY----- 405
Qy 318 AKPGSADLPENTNIATINPNTSNDPGQKVTVRDGQITIKKIDGSTKASLQAGAFVLKNA 377
Db 406 ----GADLTSTTAV-----KSTLVGGLMTKTDASTNAALAGATTVVRA 447
Qy 378 TCGQ-----PLNFND---TNVFWGTEANATE-----YTTGAD 406
Db 448 DNKEDAQDFEANAAYFNNSASGGIVTNLTSSKAAPVTGDTSGNANTSATAPVTFITGKD 507
Qy 407 GIITITGLK-----EGT----YILVEKKAPLGYNL-----LDNSQKVLGDGATDT 448
Db 508 GIATFNLGLNLDVNDGSTWTHYILVEVAAPTYQLPSVTTAANTGAVTASTAPAAITD 567
Qy 449 TNSDNLVNPTVENNKGTLPSTGGIGTTFIYIIGAILVIG--AGIVLVARRR 499
Db 568 -----TITNKKFPALPFTGGG-----LAGIIAIATVSGVIAPIAKR 604

RESULT 17
ID ABP29802
XX AC ABP29802;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 8780.
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
OS Streptococcus agalactiae.
XX WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Maignani V, Margarit Y RosI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN70433.
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX PS Claim 1; Page 3995; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
```

```
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5493 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX Sequence 554 AA;
SQ
Query Match 9.5%; Score 244.5; DB 5; Length 554;
Beat Local Similarity 25.4%; Pred. No. 1.5e-06;
Matches 151; Conservative 61; Mismatches 218; Indels 165; Gaps 31;
Qy 1 MKKKMIQSLLVASLAFGNVSPVTPPIAFAAETGTTV-----QDTQKGAIFYKAVK--- 51
Db 3 LSKKLLFSAALVTWAGSTVEPVAQFA---TGMSIVRAAEVSQERPAKTTVNIYKLOAD 58
Qy 52 -FDAEIDNANVSDSNKDG--ASYLIPQGEAEYKASTDFNSLFTTTNGGRTYTVTKDTA 108
Db 59 SYKSEI-TSNGGIENKGEVISNYAKLGDNVKLGQGVF-----KRYKVTB-I 105
Qy 109 SANEIATWAKSISANTTPVSTVTE-----SNNDGTEV---INVQYGYVYSSTVNN 157
Db 106 SVDELKK-LTTVEAADAKVGTILEEVSGLPQKTAQGLVVDALDSKSNRVLYVEDLKNS 164
Qy 158 GAVIMVTSVTP-----NAT-----IHEKN--TDATWGGGKTVQDKTYSGVD 198
Db 165 PSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKVKKLGQ-----D 218
Qy 199 TVKTYTYKNAVNVHGTGKTVYVVKDTPMPSASVVDLNEGSYE-VTITDGSNITTLTQG 257
Db 219 DAGYTI-----GEE--PKWFLKSTIPA-----NLGDYKFEITDKFADGLTY--- 258
Qy 258 SEKATGKYNL-----LEENNNFTITIPWAATNPTGNTONGANDDP-----FYKGI 303
Db 259 --KSVGKIKIGSKTLNRDEHYTI-----DEPTVDNQNTLKITFKPEKFEIABELLKGM 309
Qy 304 NTITVYTYGVLKSGAKPGSADLPENTNIATI-----NPN 337
Db 310 -TLVKNQDALDKATANTDDAAFLPIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPK 368
Qy 338 TSNDPQGVKTVRDGQITIKKIDGSTKASLQAGAFVLKNAVQGLNFNDTNNVWGTEAN 397
Db 369 PSN--PPRKPEVHTGGKRFVKDSTETQTLGSAFEDLLASDGTAVKWTDA-LIKANTKN 425
Qy 398 --ATEYTTG-----ADGIITITGL-----KEG---TYLVEKKAPLGYNLNDSQ 437
Db 426 YIAGEAVTGQPIKLKSHDGTGFEIKGLAYAVDANAEGTAVTYKLKETRAPEGVVPDKEI 485
Qy 438 KVLGDGA-----TDTTNSDNLVNPTVENNKGTLPSTGGIGTTFIYIIGAILV 487
Db 486 EFTVSQTSYNTKPTDITVDSADATPDPTIKNNKRPISPTNGGIGTAIFAIVGAAMV 540

RESULT 18
ABP27373
ID ABP27373 standard; protein; 581 AA.
XX AC ABP27373;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 3922.
```

XX	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW	group A streptococcus; Streptococcus pyogenes; antibacterial;
KW	antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX	Streptococcus agalactiae.
OS	
XX	WO200234771-A2.
PN	
XX	02-MAY-2002.
XX	29-OCT-2001; 2001WO-GB004789.
PP	
XX	27-OCT-2000; 2000GB-00026333.
PR	24-NOV-2000; 2000GB-00028727.
PR	07-MAR-2001; 2001GB-00005640.
XX	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
PA	Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
XX	Tettelin H;
PI	
XX	WPI; 2002-352536/38.
DR	N-PSDB; ABN68004.
DR	
XX	New Streptococcus protein for the treatment or prevention of infection or
PT	disease caused by Streptococcus bacteria, such as meningitis, and for
PT	detecting a compound that binds to the protein.
PT	
XX	Claim 1; Page 3546; 4525pp; English.
XX	
PS	The invention relates to a protein (ABP25413-ABP30895) from group B
CC	streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC	the specification. The proteins have antibacterial and antiinflammatory
CC	activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC	antibodies that bind (I) are used in the manufacture of medicaments for
CC	the treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a
CC	biological sample. (I) is used to determine whether a compound binds to
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (I) may be used to recombinantly produce (I) and may be
CC	used in gene therapy. Antibodies to (I) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
CC	Streptococcus proteins
XX	
SQ	Sequence 581 AA;
Query Match	9.5%; Score 244.5; DB 5; Length 581;
Best Local Similarity	25.4%; Pred. No. 1.6e-06;
Matches 151; Conservative	61; Mismatches 218; Indels 165; Gaps 31;
QY	1 MKKQMIQSLVLASLAFGMAVSPVTPPIAPAAFTGTTV-----QDTQKGYKAYKV--- 51
Db	30 LSKLLFSAALVLTWAGSTVEPVAQFA---TGMISVRAAEVSGERPAKTVNIYKLQAD 85
QY	52 -FDAIDNANVSDSKDG--ASYLIPQKEAEYKASTDFNSLFTTTTNGGRTYVYKKDTA 108
Db	86 SYKSEI-TSNGGIENKDEGISVNAKLGDNVKGLOGVQF-----KRYKVKTD-I 132
QY	109 SANETATWAKSTISANTTPVSTVTE-----SNNDGTEV---INVSQYGVYVSSVTNN 157
Db	133 SVDELKK-LTTVEARADAKVGIILEEGVSLPQKTNAAQGLVVDALDSKNVRVLYVEDLKN 191
QY	158 GAVIMVTSVTP-----NAT-----IHEKN--TDAWGDGGKTVQDKTYSVGD 198
Db	192 PSNITKAVAVPVLPELPVANSTGTGFLSEINIPKNNVTVDBPKTDKQVKLQG-----D 245
QY	199 TVKYITYTKAVNVYHGTEKVQYVIKDTMPSASVVVDLNEGSVE-VTITDGSNITTTLTQ 257

Db	246	DAGYTI-----GEE--FKWFLKSTIPA-----NLGDYKFEITDKFADGLTY--- 285
QY	258	SEKATGKYNL-----LENNFTTITIPWAAVNTPTGNTONGANDDF-----FYKGI 303
Db	286	--KSVGKIKIGSKTLNRDEHYTI-----DEPTVDNQNTLKITFKPEKFEIAELLKGM 336
QY	304	NTITVTVTGVLKSGAKPGSADLPENTNIATI-----NPN 337
Db	337	-TLVKNQDALDKATANTDDAAFLPIPVASTINEKAVLGKAIENFTFELQYDHTDPKADNPK 395
QY	338	TSNDPQGVVTRDQGITIKIDGSTRASLOGAIFVLKNATGQFLNFNDNNVWGTAN 397
Db	396	PSN--PPRKPEVHTGGKRFVKKSTETQTLGGABFDLLASDGTAVKWTDA-LIKANTKN 452
QY	398	--ATEYTTG-----ADGIITITGL-----KEG---TYLVEKKAPLGYNLLDNSQ 437
Db	453	YIAGEAVTGQPIKLKSHDTGTFEIKGLAYADANAEGTAVTYKLYKAPGYVDPKEI 512
QY	438	KVILGDGA-----TDTTNSDNLVNPTVENNKGTELPSTGGTGTTFIIGAILV 487
Db	513	EFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPTNTGGTGTGTAIFVAIGAAYM 567

RESULT 19  
ADC95757  
ID ADC95757 standard; protein; 341 AA.  
XX  
AC ADC95757;  
XX  
DT 01-JAN-2004 (first entry)  
XX  
DE E. faecium protein sequence SEQ ID 5384.  
XX  
KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
XX abdominal-pelvic infection.  
XX  
OS Enterococcus faecium.  
XX  
PN US6583275-B1.  
XX  
PD 24-JUN-2003.  
XX  
PF 30-JUN-1998; 98US-00107532.  
XX  
PR 02-JUL-1997; 97US-0051571P.  
PR 14-MAY-1996; 98US-0085598P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Doucette-Stamm LA, Bush D;  
PI  
XX  
XX WPI; 2003-799836/75.  
DR  
N-PSDB; ADC92103.  
XX  
XX New isolated nucleic acid derived from Enterococcus faecium encoding an  
PT Enterococcus faecium polypeptide useful for detection, prevention and  
PT treatment of a pathological condition resulting from a bacterial  
PT infection.  
XX  
XX Example 1; SEQ ID NO 5384; 243pp; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid derived from  
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
CC one of 10 fully defined sequences given in the (or comprising 40  
CC sequential nucleotides chosen from any of the nucleic acids, its  
CC complement or sequences hybridising to it). Also included are a  
CC recombinant vector comprising the nucleic acid operably linked to  
CC transcription regulatory element, a cell comprising the vector and a  
CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
CC The nucleic acids is useful for diagnosing pathological conditions  
CC resulting from E. faecium bacterial infection (e.g. urinary tract

CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
CC (infection) and for screening drugs such as agonists and antagonists. The  
CC nucleic acid is useful for recombinant production of *Candida albicans* -  
CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
CC and vaccines containing the nucleic acid are useful for preventing or  
CC treating *Enterococcus faecium* infections. The present sequence represents  
CC one of the disclosed *E. faecium* proteins.  
XX  
SQ Sequence 341 AA;

Query Match 9.1%; Score 234.5; DB 7; Length 341;  
Best Local Similarity 28.5%; Pred. No. 3.2e-06;  
Matches 96; Conservative 44; Mismatches 138; Indels 59; Gaps 16;

QY 190 DQKTVSV-----GDTVKYIT-----YKNVNHGTEKYQYVIKDTWPSASVVDLNEGS 239  
DB 15 EQETISYDLRGKTASYITAPIPYFIDSVLNGSAVIKNYKITDT---PTVGLTYVD 70  
QY 240 YEVITDQSGNITLTQSEKATGKYNLLENNNTIIPWAATPTGNTONGANDDF 299  
DB 71 QIEVRAGE---TILTKGQD-----YIVEVWSNGFVVTILTBEENGVAKYVDLTLRLAD--- 119  
QY 300 YGINTITVYTGVLKSGAKPGSADLPENTNATINPNTSND-----DPGQKTVYRDGQI 354  
DB 120 ARG-GDLHITYN--LKVSTELADDFHNTAVIEGRNDEFYERGVEPEKVT--GGR 174  
QY 355 TIKKIDGSKASLOGAIFVLKNA-TGQFLNF-----NDTNVWVGTEANATBYT 402  
DB 175 KPEKYDASSSELKQARPELWNEDESRVAIFYKGSPLAVYSGADRIEAWTSGQATEFV 234  
QY 403 TGADGIITITLKEGTYYLVEKAPLGNLLNS---QKVLGDGATDTTNSDNLV--- 456  
DB 235 ADGNGYFEQGLDYGTYQMKETMAPEGYVLPTEGAATFEFIISYG---SYNBEIQIVGVE 291  
QY 457 NP---TVENNKTELPSTGGIGTPTIPIIGAILVIGA 490  
DB 292 NPGPERVPMKRGSLPATGGNGLLAFLIGISLMIGA 328

RESULT 20  
ID ABB47318 standard; protein; 793 AA.

XX ABB47318;

AC ABB47318;

DT 05-FEB-2002 (first entry)

DE *Listeria* monocytes protein #22.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KW vitamin B12; bacterial infection; disease.

XX *Listeria* monocytes.

XX WO200177335-A2.

PN 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR001118.

PF 11-APR-2000; 2000FR-00004629.

XX (INSP ) INST PASTEUR.

PA Buchreier C, Frangeul L, Couve E, Ruaniock C, Psihi H, Dehoux P;

PI Dusurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;

PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

PI Rose M, Voss H;

XX

DR WPI; 2002-010914/01.

XX Genomic sequence for *Listeria* monocytes, useful e.g. for treatment  
PT and prevention of *Listeria* and related bacterial infections, and related  
PT polypeptides.

PS Claim 6; SEQ ID NO 23; 192pp; French.

XX The present invention relates to the genome sequence of *Listeria*  
CC monocytes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in *L.*  
CC monocytes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of *L.* monocytes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate *L.* monocytes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by *L.*  
CC monocytes and related organisms. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 793 AA;

Query Match 9.0%; Score 232; DB 5; Length 793;

Best Local Similarity 25.1%; Pred. No. 1.3e-05;  
Matches 125; Conservative 66; Mismatches 189; Indels 118; Gaps 23;

QY 45 TYKAYKVF--DAEIDNANVSDNKGASYLIPQ-----GKEAEYK---- 82  
DB 239 TYSADRVFTAGAPISSTNFS-ATSDGFSVALGNLTDVQISYITTTTGGKSTQDNTAK 297  
QY 83 -ASTDF-----NSLFTTTTNG-----GRYTVTKD--TASANEIATW-----A 117  
DB 298 LAGTDFVTQKSTWTPASGGGEGGTGTVLTKEADAKTKATLEGAEFLVDKSGTVLQ 357  
QY 118 KSISANTPTPVSTVESNNDGTEVINVSQYGYVYSST-----VNGCAVIMVTSVTPNA 170  
DB 358 ENITTNASGQUSIADLPDFTYOLIEETKAPTGYKLDTPVEFTIGENNOAI----TVTKEN 413  
QY 171 TIHEKNTDATWGDGCGK-TVDQKTVSGDVTVKYITTYKNAVNVHTEKV-----YQVY 222  
DB 414 TLNTGVELTKLDAATKATLAGATFELQDKEGNTLQDLDKTDENGVLKVDLVPGSYQFV 473  
QY 223 IKDTWPSASVVDLNEGSYEVTITDQSGNITLTQSEKATGKYNLLENNNTIIPWAA 282  
DB 474 -ETSAPTGYKLDNSPVSEV-IAGTDQVVKVT-----KENTLEVG-----SVELTK 518  
QY 283 TMTPTGNTQANDDDFFYKGINITITVTV-----TGVLK-SGAKPGSADIPEN---TNIATI 334  
DB 519 LDSATKATLAGATFELQDKEGNTLQDLDKTDENGVLKVDLVPGSYQFVTSAPTGYKLD 578  
QY 335 NPNTS-----NDPQGVKTVRD-----GOITKIKIDGSKASLOGAIFVLKNAATQFLNF 384  
DB 579 NSPVSEVAVAGETQVVKVTENTLEVGSEVLTCLDSATKATLAGATFELQDKEGNTLQ- 637  
QY 385 NDTNNVWVGTEANATEYTTGADGIITITGLKEGTYYLVEKAPLGNLLNSDKYILGDG 444  
DB 638 -----TGLTTDENGVLKVDLVPGTQVQFVETKAPIGYELDTTPVSPFVAVG 683  
QY 445 ATD---TTNSDNLVNPT 459  
DB 684 ETDPIVKTKENTLVPPT 701

RESULT 21

ABU32765

ID ABU32765 standard; protein; 793 AA.



Matches	125;	Conservative	66;	Mismatches	189;	Indels	118;	Gaps	23;
QY	45	TYKAYKVP--DAEIDNANVSDSKDQASYLIPQ-----GKEAEVK-----	82						
DB	239	YSISADRVFTAGAPISSTNFG-ATSDGFSVALGNLDSVOISVTTTTTGGKSTQIDNTAK	297						
QY	83	-ASTDF-----NSLFTTTTNG-----GRVTYTKKD--TASANEIATW-----	117						
DB	298	LACTDFTVTKQTSITWTPASGGGEGGGCTTGSVTLTKREDAKTATLEGAEFLVDSKGTVLQ	357						
QY	118	KSISANTTPVSTVTESNNDDTEVINVSQYVYVSGT-----VNGAVIMVTSVTENA	170						
DB	358	ENIITNASGQLSIADLKFDYQLIETKAPTGYKLDTPVEFTIGENNOAI-----TVTKEN	413						
QY	171	TIHEKWTDAITWGGGGK-TVDQKTSYVGDTVKYTIYKNAVNVHGTGV-----YQYV	222						
DB	414	TLNTGSVELTKLDAATKATLAGATFELQDKEGNLTQDLKTDENGVLKVTDLVPGSYQFV	473						
QY	223	IKDWTMPSASVVDLNEGSEYVTTIDGSGNITTLTQGSSEKATGKYNLLENNNNFTIIPWAA	288						
DB	474	-ETSAPTGYKLDNSPVSFEV-IAGETDQVVKVT-----KENTLEV-----SVELTK	518						
QY	283	TNTPGTNTQNGANDFPYKGINITVY-----TGVLK-SGAKPGSADLPEN---TNIATI	334						
DB	519	LDSATKATLAGATFELQDKEGNLTQDLKTDENGVLKVTDLVPGSYQFVTSAPTGYKLD	578						
QY	335	NENTS-----NDDPGOKVTRD---GQITIKKIDGSTKASIQGAIFVLKNATQOFLNF	384						
DB	579	NSPVSFEVAGETDQVVKVTKENTLEVSGSVELTKLDSATKATLAGATFELQDKEGNLTQ-	637						
QY	385	NDNTNVWGTTEANATEYTTGADGIIITIGLKEGTYLVEKKAPLGNLLDNSQKVLGDG	444						
DB	638	-----TGLTTDENGVLKVTDLVPGTYQFVETKAPIGYELDTPVPSFEIVAG	683						
QY	445	ATD---TTNSDNLAVNPT	459						
DB	684	ETDPIVKVTKENTILPPT	701						
RESULT	22								
ID	AAM16315								
XX	AAM16315	standard; protein; 688 AA.							
AC	AAM16315;								
DT	12-OCT-2001	(first entry)							
DE	Peptide #2749	encoded by probe for measuring cervical gene expression.							
KW	Probe; human; microarray;	gene expression; cervical epithelial cell;							
KX	cervical cancer.								
OS	Homo sapiens.								
FN	W0200157278-A2.								
PD	09-AUG-2001.								
PF	30-JAN-2001; 2001WO-US0000670.								
PR	04-FEB-2000; 2000US-01803112P.								
PR	26-MAY-2000; 2000US-0207456P.								
PR	30-JUN-2000; 2000US-00608406.								
PR	03-AUG-2000; 2000US-00632366.								
PR	21-SEP-2000; 2000US-0234687P.								
PR	27-SEP-2000; 2000US-0236359P.								
PR	04-OCT-2000; 2000GB-00024263.								
XX	(MOLE-) MOLECULAR DYNAMICS INC.								
XX	Penn SG, Hanzel DK, Chen W, Rank DR;								
XX	WPI; 2001-488901/53.								



XX AAM28810;  
AC 17-OCT-2001 (first entry)  
DT  
DE Peptide #2847 encoded by probe for measuring placental gene expression.  
XX  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200157272-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US000663.  
XX  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488897/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human placenta.  
XX  
XX Claim 27; SEQ ID NO 29079; 654pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SNP:  
XX see AA131315-AA157546). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for producing a microarray for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from human placenta. The probes are useful for antenatal diagnosis of  
XX human genetic disorders  
XX  
XX SQ Sequence 688 AA;  
  
Query Match 8.6%; Score 222.5; DB 4; Length 688;  
Best Local Similarity 24.3%; Pred. No. 4.1e-05;  
Matches 115; Conservative 57; Mismatches 221; Indels 81; Gaps 18;  
  
QY 32 TGTITVQDTQKATYKAYKVFDAEIDNANVSDNKGASLYLPQKAEAYKASTDFNSLF 91  
DB 38 TGGTATKVTGSETTTTSTEGSEITATISITGETTTAS--TEGSETT-TASTEGSETT 93  
QY 92 TTTTNGRTYVKKDTASANEIA-----TWAKSISANTPVTSTVSTESNNDGTEVINVS 144  
DB 94 SASITGSET-TTASTTSSETTASIMGSETTMASTIGSETTKVST---ASRKMVTTF--- 146  
QY 145 QYGYVYSSVTVNGAVLWTVSTPNATIEHKNTDATWGDGGKTVDQKTVSGDVTVKYTI 204  
DB 147 -----TENSETTASTASTSTVSTAGSETTIPASTAGSETTTTSTEGSEITATIS 197  
QY 205 TYKNVANYHGTEKYVYVVIDTWPESASVVDLNEGSEYVVI-TDQSGNITTLTGSEKATG 263  
DB 198 TEGSEITTTASTES-----SETTTATTI-----GSETTTASTEGSEITTTSTEGSETTTA 246  
QY 264 KYNLENNNFTTIPWAATNTPGNTONGANDFFPKGINTTIVTGTGVLKSGAKPGSA 323  
DB 247 S-----TEGSEIT-TVSTTGSTTTTASTEGSEITTTASTEGSELTVSTGTG---SETITVSA 298  
QY 324 DLPTNTIATINPNTSNDPDG-----QKVTVRDGQITIKKIDGS--TKASLQGAIFVLKNA 377  
DB 299 EGSETTIVTMGSETTTASTAGSETTTVSTAGSETTTASTAGSETTTVSTGSETTTVST 358

QY 378 TQOFLNFNDNNVW-----GTEANAFETYTTGADGIITITLKGTYVLVEKKAPLGYN 431  
DB 359 TGTETITSTEGSEITTTVTTAGSETTAV-YTTGSETTTTSTEGSETT----- 404  
QY 432 LLDNSQKVLGSGADTTNNSDNLVNPTVE--NNKGTELPSTGGTGTTTFYIIGA 484  
DB 405 -----TVSTTGSTTTTASTADLETTTVSTSGSGTTTTASTAGSETTTVYITGS 451  
  
RESULT 25  
ABB30137  
ID ABB30137 standard; peptide; 688 AA.  
XX  
XX AC ABB30137;  
XX  
XX DT 01-FEB-2002 (first entry)  
XX  
XX DE Peptide #2788 encoded by breast cell single exon nucleic acid probe.  
XX  
XX KW Human; microarray; single exon probe; gene expression; breast; disease;  
XX cancer.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157271-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US000662.  
XX  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-496933/54.  
XX  
XX New spatially-addressable set of single exon nucleic acid probes, useful  
XX for measuring gene expression in sample derived from human breast,  
XX comprises number of single exon nucleic acid probes.  
XX  
XX Claim 27; SEQ ID NO 13105; 327pp + Sequence Listing; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human breast and BT 474 cells. The method involves contacting the  
XX probes with a collection of detectably labelled nucleic acids derived  
XX from mRNA of human breast, and then measuring the label bound to each  
XX probe of the microarray. The probes are useful for verifying the  
XX expression of regions of genomic DNA predicted to encode proteins. They  
XX are useful for gene discovery, and for determining predisposition and/or  
XX prognosing breast disease. Gene expression analysis is useful for  
XX assessing the toxicity of chemical agents on cells. The microarray of  
XX this invention presents a far greater diversity of probes for measuring  
XX gene expression, with far less bias than expressed sequence tag  
XX microarrays. The method is suitable for rapid production of functional  
XX information from genomic sequence. The present sequence is a peptide  
XX encoded by a single exon nucleic acid probe of the invention. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 688 AA;  
  
Query Match 8.6%; Score 222.5; DB 4; Length 688;

Best Local Similarity 24.3%; Pred. No. 4.1e-05;  
Matches 115; Conservative 57; Mismatches 221; Indels 81; Gaps 18;

Qy 32 TGITTVDTQKGATYKAYKVDFDAIDNANVSDSNKDGSYLIPQCKEAEYKASTDENSELF 91  
| | : : : : :  
Db 38 TGSEITKVSTGSETTTTSGSEIITASIGSETTAS---TGSETT-TASTEGETT 93  
| | : : : : :  
Qy 92 TTTTNGRTYVTKDTSANEIA-----TWAKSISANTPPSVTSNNNDGTVEINVVS 144  
| | : : : : :  
Db 94 SASSTGSET-TTASTTSSETTMASIMGSETTMASTIGSETTKVST--ASSKMVTVF--- 146  
| | : : : : :  
Qy 145 QYGVVVSSTVNNGAVIMVSVTPNAIHKNVTDATWCDGGGKTVDQKTVSGDTVVKYTI 204  
| | : : : : :  
Db 147 -----TENSETTIASITASETTTIVSTAGSETIPASTAGSETTTTTTTEGSETTTAS 197  
| | : : : : :  
Qy 205 TYKANVNHGTEKYVYVIKDTMPSASVVDLNEGSYEVTI-TDGSGNIITLTQSEKATG 263  
| | : : : : :  
Db 198 TEGSETTTASTES-----SETTTATII-----GSETTTASTEGETTTTSTEGETTTA 246  
| | : : : : :  
Qy 264 KYNLLENNFTIPIPAWANPTGNTONGANDFFYKGINITIIVTYGVLVLSGAKPESA 323  
| | : : : : :  
Db 247 S----TEGSEIT-TVSTTGSTTTASTEGSETTTASTEGSELTTVSTTG---SETITVSA 298  
| | : : : : :  
Qy 324 DLPENTMIATNPMTSNDPG-----QKVTVRDQGITIKKIDGS--TKASLOGAIFVLKNA 377  
| | : : : : :  
Db 299 EGSETTTVTMGSETTTASTAGSETTTVSTAGSETTTASIEGSETTTVSVSTGSETTTVST 358  
| | : : : : :  
Qy 378 TGQFLNFENDNNVEW-----GTEANATEYTTGADGIITITGLKEGYLVLEKKAPLGYN 431  
| | : : : : :  
Db 359 TGTHETTTTSGESTTTTVTAGSETTAV-YTIGSETTTTTEGSETT----- 404  
| | : : : : :  
Qy 432 LLONSQKVLGDGADTTNSDNLLVNPTVE-NNKGTPLPSTGGIGTTFIFYIGA 484  
| | : : : : :  
Db 405 -----TVSTTGSTTTASTADLETTVSTVSGSGTTTASTAGSETTTVYTIGS 451  
| | : : : : :

RESULT 26  
ABB20749  
ID ID ABB20749 standard; protein; 688 AA.  
XX AC ABB20749;  
XX AC  
XX DT  
XX DT  
XX DE  
XX DE  
XX KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX OS Homo sapiens.  
XX PN WO200157274-A2.  
XX PD  
XX PF 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US0000666.  
XX PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00612366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PA  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR  
XX DR WPI; 2001-488899/53.  
XX PT Single exon nucleic acid probes for analyzing gene expression in human

PT	hearts.
PS	Claim 15; SEQ ID NO 22519; 530pp; English.
XX	The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21335-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 688 AA;
XX	
XX	Query Match            8.6%; Score 222.5; DB 4; Length 688;
XX	Best Local Similarity 24.3%; Pred. No. 4.1e-05;
XX	Matches 115; Conservative 57; Mismatches 221; Indels 81; Gaps 18
Qy	32 TGTITVQDTQGATYKAYKVFAEIDNANVSNSKDGYSLIPOKEAEYKASTDFNSLIF 91     : : : : :     : : : : :     : : : : :     : : : : :
Db	38 TGSETTKVSTTGSETTTTTSGSEITTASITGSETTTA--TEGETT-TASTEGETT 93     : : : : :     : : : : :     : : : : :     : : : : :
Qy	92 TTTTNGGRYYVKDXTASANEA-----TWAKSISANTTPVSVTVTESNNDGTEVINVS 144 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	94 SASITGSET-TTASTTSETTMASIMGSETTMAGTIGSETTKVST---ASSIKMTTFV---- 146 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	145 QYGVVVSSSTVNGGAVIMVTSPNATTHEKNTDATMGDGKGKTVDQKTYSVGDIVKYTI 204 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	147 -----TENSETTIASTTASETTTIVTAGSETIPASTAGSETTTTSTEGSETTTAS 197 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	205 TYKNADVNHGEKYYVIKDMPSASVVDLNESYEVTI-TDGSNNITTLTQSGEKATG 263   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	198 TEGSETTTASTPES-----SETTTATTI-----GSETTTASTEGSETTTTSTEGSETTTA 246   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	264 KYNLEENNNFTIIPWAANTPGNTONGANDFFYKGINITIVTYGLSKGAKPGSA 323   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	247 S-----TEGSEIT-TVSTTGSETTTASTEGSETTTASTEGSELTTVSTTG---SETITVSA 298   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	324 DLPENTNIATNPNTSNDDPG---QKVTVRDGOITIKKIDGS--TKASLOGAIFVLKVA 377   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	299 EGSETTTVTMGSETTTASTAGSETTTVSTAGSETTTASIEGSETTTVSTGSETTTVST 358   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	378 TGQFLNFNDTNVNW-----GTEANAPEYTTGADGIITITGLKEGTYYLVVEKKAPLGYN 431   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	359 TGTETTTTSTEGSETTTVTAGSETTA-VYTTGSETTTTSTEGSETT----- 404   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	432 LLDNSQVILGDGATDTNSDNLLNPWVE-NNKGTPLSPGGIGTIFIYIGA 484 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	405 -----TVSTTGSETTTASTADLETTVSTSGSGTTTASTAGSETTTVYTGS 451 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
XX	
XX	RESULT 27
XX	AAM68511
ID	AAM68511 standard; protein; 688 AA.
XX	AC AAM68511;
XX	DT 06-NOV-2001 (first entry)
XX	DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28817.
XX	KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX	KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX	OS Homo sapiens.
XX	FN WO200157276-A2.
XX	PX

# PT Single exon nucleic acid probes for analyzing gene expression in human

[illegible]



Search completed: July 22, 2004, 10:27:32  
Job time : 60 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 10:21:56 ; Search time 19 Seconds  
(without alignments)  
2541.481 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKMKIQLLVASLAFGVAV.....GALLVIGAGIVLVARRLRS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

PIR 78:\*

1: Pirl:\*

2: Pirl2:\*

3: Pirl3:\*

4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	284	11.0	537	2 A35400	surface protein T6
2	265.5	10.3	665	2 F95053	cell wall surface
3	258.5	10.0	614	2 F86719	hypothetical prote
4	258	10.0	533	2 A35259	type 1 fimbrial pr
5	251	9.7	507	2 S52348	hypothetical prote
6	232	9.0	793	2 AH1094	probable peptidogl
7	231.5	9.0	534	2 A32347	fimbrial protein t
8	219	8.5	1612	2 AB1347	probable peptidogl
9	217	8.4	562	2 AE1335	probable peptidogl
10	216	8.4	1806	2 AF1717	probable peptidogl
11	214.5	8.3	881	2 AC1458	surface anchored p
12	212	8.2	2551	2 F98047	hypothetical prote
13	205	8.0	1622	2 AE1717	probable cell surf
14	200.5	7.8	893	2 E95053	cell wall surface
15	200	7.8	1441	2 B86807	hypothetical prote
16	199	7.7	1530	2 AH1396	peptidoglycan anch
17	192.5	7.5	571	2 AI1094	probable peptidogl
18	191	7.4	1711	2 AB1283	peptidoglycan link
19	189	7.3	1519	2 S41525	major ring-forming
20	186.5	7.2	1275	2 T33369	hypothetical prote
21	186	7.2	3705	2 AD0123	probable autotrans
22	185.5	7.2	583	2 S57721	cspB protein - Clo
23	185.5	7.2	1026	2 A48995	atacrystalline su
24	185.5	7.2	1073	2 C87374	S-layer protein R8
25	184.5	7.2	2468	2 AB3412	hypothetical prote
26	181.5	7.0	1315	2 T28679	fibrinogen-binding
27	181	7.0	3029	2 S76109	hypothetical prote
28	180	7.0	1166	2 T28680	fibrinogen-binding
29	180	7.0	1231	2 T28681	rib protein - Stre

#### RESULT 1

A35400

Surface protein T6 precursor (strain D471) - Streptococcus pyogenes

C:Species: Streptococcus pyogenes

C:Date: 31-Aug-1990 #sequence\_revision 31-Aug-1990 #text\_change 15-Oct-1999

C:Accession: A35400

R:Schneewind, O.; Jones, K.F.; Fischetti, V.A.

#### ALIGNMENTS

30	180	7.0	1286	2 S28634	adhesin AIDA-I pre
31	179.5	7.0	2021	2 A97859	190-KDa cell surfa
32	178.5	6.9	4936	2 AH2515	hypothetical prote
33	178	6.9	1041	2 S55862	probable membrane
34	177.5	6.9	1185	2 A42404	collagen adhesin -
35	177.5	6.9	1965	2 S75200	fat protein - Syne
36	177	6.9	1029	2 T30852	outer membrane pro
37	176.5	6.8	1385	2 D89824	hypothetical prote
38	176	6.8	1276	2 B86546	polymorphic outer
39	176	6.8	1276	2 C81591	polymorphic membra
40	175	6.8	1536	2 A43855	high-molecular-wei
41	175	6.8	1999	2 AB2018	hypothetical prote
42	174.5	6.8	1461	2 E90696	hypothetical prote
43	174.5	6.8	1461	2 B85547	hypothetical prote
44	174.5	6.8	5188	2 B85547	probable RTX fami
45	174.5	6.8	5291	2 F90696	hypothetical prote
46	174	6.7	1428	2 AC2224	hypothetical prote
47	173.5	6.7	586	2 AD1458	probable peptidogl
48	173	6.7	359	2 S42787	serine/threonine-r
49	173	6.7	1528	2 D85912	hypothetical prote
50	173	6.7	1571	2 C91068	hypothetical prote
51	172.5	6.7	393	2 G95053	cell wall surface
52	172	6.7	953	2 C89824	hypothetical prote
53	171.5	6.7	1407	2 B72078	polymorphic outer
54	171.5	6.7	1983	2 G86643	hypothetical prote
55	171.5	6.7	2554	2 AB3528	extracellular seri
56	171	6.6	595	2 B48658	flagellin - Bacher
57	171	6.6	1661	2 A83695	hypothetical prote
58	169.5	6.6	1348	2 AH1115	cell surface prote
59	169	6.6	1873	2 T30944	surface protein pr
60	169	6.6	2249	2 A41477	190K surface antig
61	168	6.5	821	2 AD1507	probable secreted
62	168	6.5	1910	2 AF0394	probable adhesin h
63	168	6.5	2044	2 AB1180	probable peptidogl
64	168	6.5	3013	2 AB0480	probable invasin y
65	168	6.5	3083	2 AH2493	hypothetical prote
66	167.5	6.5	414	2 AG1459	probable peptidogl
67	167.5	6.5	826	2 AB1841	hypothetical prote
68	167.5	6.5	939	2 I41197	eae protein (enter
69	167.5	6.5	1902	2 B45764	lactococpin (EC 3.4
70	167	6.5	1141	2 E89824	hypothetical prote
71	167	6.5	2020	2 C48399	ABC-type transport
72	166	6.4	821	2 AD1148	probable secreted
73	166	6.4	940	2 AB1744	internalin protein
74	165.5	6.4	949	2 D90803	Aida-I adhesin-lik
75	165.5	6.4	1005	2 H85611	probable adhesin Z
76	165.5	6.4	1588	2 A86036	probable adhesin Z
77	165.5	6.4	1588	2 H91188	probable adhesin E
78	165.5	6.4	1902	2 S06997	lactococpin (EC 3.4
79	165	6.4	1032	2 T34433	hypothetical prote
80	165	6.4	1386	2 AC1533	surface protein (L
81	165	6.4	1643	2 D71630	outer membrane pro
82	165	6.4	1655	2 E97835	hypothetical prote
83	165	6.4	1816	2 F83901	hypothetical prote
84	165	6.4	2232	2 T34434	hypothetical prote
85	164.5	6.4	794	2 T36972	probable membrane
86	164.5	6.4	1268	2 B99789	hemagglutinin/hemo
87	164.5	6.4	1270	2 E85649	hypothetical prote
88	164.5	6.4	1962	2 A32634	lactococpin (EC 3.4
89	164	6.4	961	2 AD0548	putative autotransp
90	164	6.4	1569	2 A85044	hypothetical prote

J. Bacteriol. 172, 3310-3317, 1990  
A; Title: Sequence and structural characteristics of the trypsin-resistant T6 surface protein  
A; Reference number: A35400; MUID: 90264329; PMID: 2188957  
A; Accession: A35400  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-537 <SCH>  
A; Cross-references: GB:M32978; NID:G153842; PIDN:AAA27019.1; PID:G153843  
C; Keywords: transmembrane protein

Query Match	11.0%	Score 284;	DB 2;	Length 537;
Best Local Similarity	22.2%;	Pred. NO. 3.7e-08;		
Matches 135: Conservative	89;	Mismatches 195;	Indels 188;	Gaps 28;

```

Qy 10 LVASLAF-----GMAVSPVTPIAFAAETGITTVODTQGATYKAYKYVPDAEIDNANVSDSN 65
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 1 MLACLALAVVGCGMTVRVSALS-KDDTAQLUKITNIEGGTPVTLYKIGEGVY-----N 51
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 66 KDGASYLIPOGKEAEYKASTD---FNSLFTTTNGGRYYTK-----KDTASANEIATW-A 117
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 52 TNGDSFI--NFKAEGVSLTGTGTSQEIIITIANINTGKIKPSTENUSISNGTATYNA 109
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 118 KSISA-----NTTPSVTVTESNNNDGTEVI-NV-SOQYGVYVSSTVNAGAVIMV 163
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 110 RGASVYALLTGATDRGTYNPILLAASYNGEGLVTKNIDSKNLYLYGOTSVAKSSLPISI 169
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 164 TSUPTNATIHEKNTDATMGDCGGKTVDOKTYSVGDTVKYTIIV-----KNAYNVHGTEK 217
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 170 T-----KKVTGTIDD-----VNKKTSLGSVLSTLTPELPSYTKEANV----KT 210
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 218 VYQYVIDTMPASVVDNLNEGSVEVTTITDGSIGNITTLTOGSEKATGKYNLLLENNNPFIT 277
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 211 VY---VSDNMSEGLTFENF---SLTV-EMKGKMANITEDG-----SYMVENTKIGI- 254
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 278 IPWAATNPTGNTONGANDOFFYKIGINIIT--VTTYTGVLKSGAKPGSADIPENTN-IATI 334
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 255 -----AKEVANGFNLSFYDLSLESISPNI SYKAVANNKAIVGBEGPNPKABFFYSN 305
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 335 NPNTSN-----DDPGQKVTVRDG-----QITIKKIDGSTKASLOGAIFVLKNATG 379
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 306 NPTKGNTYDNLDLRPKDNGINGITSKEDSKIYVYQIAFRKVDVSVKTPJIGAIFGVYDTSN 365
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 380 QFLNFDNPNVWEWGTEANATEYTTGAGCIITITGLKEGTYTVLVEKKAPLGNL----- 432
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 366 KLIDI-----VTTNKGYAIATQVSSGKYKIKELKAPKGYSLNTETEYI 409
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 433 -----LDINSQKV----- 439
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 410 TANMVTAFTVKSANSKSTYTTSYDKRKNATDNSEQVWLKNGIFYSIDSRPTGNDVKEAYIE 469
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 440 ---ILGDGAT---DITTNSDNLLVNPVENNKGTELPTGGIGITTFVIIIGAILVIGA-GI 492
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 470 STKALTDTGTTFSKNEGSGTVLLETDPNTKLGEPLPTGSI GTYLFRAISGAAMIGAIGI 529
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Qy 493 VLVARRR 499
Db   :|::||:
Qy 530 YIVKERRK 536
Db   :|::||:

```

RESULT 2  
F95053  
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain T)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: F95053  
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Lofcus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95053

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-865 <KUR>  
A:Cross-references: GB:AE005672; PTDN:AAK74623.1; PTD:gl4971934; CSPDB:GN00164; TIGR:SP48  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0463

Query Match	10.3%	Score 265.5;	DB 2;	Length 665;
Best Local Similarity	24.4%;	Pred. No. 4.8e-07;		
Matches 164;	Conservative	63;	Mismatches 227;	Indels 219;
				Gaps 33;

Qy	9	LLVASLAFGNVSPVTPIAFAAETGIIITVQDTOKGATYKAYKVPDAEIDNANVSDSNKCG	68
Db	15	LLTASSLFSAA-----VFAAGTTTTSVTHKLLATDGDMDKIANELETGNTYA-GNKVG	67
Qy	69	ASYLIPQCKAEYKASDFFNSLFTTTN-----GGRTYVTTKDTASANEIATWAKGISAN	123
Db	68	---VLP---ANAKEIAGVMFVNTNNEIIDENGQTLGVNIDP-----QTFKLSGAMP	114
Qy	124	TTPVSTVTESNNDGTEVINV--SOYGYVYVSS-----TVNNGAVIMVTSVTP-----	168
Db	115	ATAMKLTAEAGAKFNTANLPAAYKIYIEHLSLSTVYGEDGATLTSKAVPIEBELPLND	174
Qy	169	--NATIEHKNTDA---TWGDDGGKT-----VDQKT--YSGVDGTVKYTTIYK--NAVNY	212
Db	175	VVDARVYPKNTEAKPKIDKDFKKGANPDPTRVDKDTFVNHOVDVEYEIVTKIPALANY	234
Qy	213	-----HGTEKYV-----QYVIKDT-----MPSASVVDLNEG--	238
Db	235	ATANWSDRMTEGLAFNKGTVKVTVDVDALEAGDYALTEVATGFDLKLTDAGLAKVNDQNA	294
Qy	239	-----SYEVTIIDGS-----GNITLTTGSEKATG---KYNLENNNFTTIPWA-A	282
Db	295	EKTWKIYTSATLNDKALIVEPESNDVTFNTGNPDHGNTPAPKNPENGDLTLTKTWVDA	354
Qy	283	TNTPTG-----NTQNGANDDFYKGINTI-----TVTYTGVLKS-----G	317
Db	355	TGAFIPAGAEATFDLVNAQTG-----KVQIVTLTTDKNTVTANGLDKNTYKFVERS	407
Qy	318	AKPOSADLPENTNIATI-----NPNTSNDDPQCKVTVRDOQIITIKKIDGSTKASLOGAI	371
Db	408	IKGYSADYQEIATTAGEIAVRNWKDENPKLPDTPTEPKVYTKGKPFVVKNDKDNR--LAGAE	465
Qy	372	FVLKNA--TGOFL-----	382
Db	466	FVIANADNAGQYLARKADKYQBEEKQLVVTIKDALDRAVAAYNALTAQQQTQOEKEVKDK	525
Qy	383	---NPN-----DTNNVBNWGTB--ANATEYTTGADGIIITITGLKEGYTYLVEKKAPLYGN	431
Db	526	AQAAYNAAVIAANNAFWEVADKONENVVKLVSDAQGRFEITGLLAGTYLYEETKQPAGYA	585
Qy	432	LLDNSOKV-----ILGDGATDTTNSDLLVNPTVENNKGTELPSTGGIGTTFIVIG	483
Db	586	LLTSRQKPEVTAISYATGQGIETAGSGK--DDATKVNNKKITIPOTGGIGTIIFAVAG	643
Qy	484	AAILVIGAGILVA	496
Db	644	AAIM---GIAVYA	653

RESULT 3  
F86719  
hypothetical protein yHge [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C.Species: Lactococcus lactis subsp. lactis  
C.Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C.Accession: F86719  
R.Polotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A.Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp  
A.Reference number: AB6625; MUID:121335186; PMID:11337471  
A.Accession: F86719  
A.Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-614 <STO>  
A:Cross-references: GB:AE005176; PID:g12723675; PIDN:AAK04856.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: yhgE

Query Match 10.0%; Score 258.5; DB 2; Length 614;  
Best Local Similarity 24.5%; Pred. No. 1e-06;  
Matches 145; Conservative 57; Mismatches 184; Indels 207; Gaps 31;

```
QY 51 VFDAEIDN-----ANVSDSNKDGASYLIPOGKEAYKASTDFNSLFTTTTNGGRTYVT 103
DB VDSGLIANNQTGGAGNTGKNDGSOQTLEGGSTPSTWANTFSATKVGTVGPTGVT 133
QY 104 K---KDTASANIATWAKSISANT-----TPVSTVTESNNDGTEVINVSQ----- 145
DB 134 DFSFTDTTSVAPVTTDASGLADFTGLTDGYILPHQVTVTVNGITVGDGFIVQVSHEDSQAG 193
QY 146 ----YGYVSVSTVNVGAVIMVTSVT-----PNATHE-----KNTD-- 178
DB 194 IVNVYPKLDMSSSAGLG-----TSATTNADDNFNGQTPNQIANPNATGNSDQTLTNTDNN 248
QY 179 ----ATWGDGGGKTVDQKTSYSGVDTVKYTIYKNAV-----NVHGTQKVV-QYV 222
DB 249 AGNEMLANGTWINGSDNQ-NTTAAAGNTVNVN---NTVFDSSQTNNGTGTGTGYII 304
QY 223 IKDTMP-----SASVVDLNEGS-----YEVTTIDGSGNI--TTLTQG 257
DB 305 VTDLQPNLNVSSVTVSTVIUNVINGSGTKVGLTPTTDYIT-NDGNGKIVVTLTTAG 363
QY 258 SEKATGKYNLLENNTFTIIPWATNTPTGNTQNGANDDFYKGINITVYTGVLKSG 317
DB 364 QOHAASL--LGSADGALNIIP-STVKSAGSATDSAT-----TTITNAY----- 405
QY 318 AKPGSADLPENTNIATINENTSDPGQKTVRVDGQITIKKIDGSKASLQCAIFVLKNA 377
DB 406 ----GADLSTTAV-----KSTLVGGLEMTKTDASTNAALAGATFTVWRA 447
QY 378 TQO-----FLFND-----TNNVEWGTANATE-----YTTGAD 406
DB 448 DNKEDAQDFVEANAAYFNNSAGGVTVNLTSSKAAFTVGTDSGNANTSATAPVTTTQGD 507
QY 407 GIITITGLK-----EGT-----YVLEKCAPGYNL-----LDNSQKVLGSGATDT 448
DB 508 GIATENGILNLDVNDGNTTNYLVVEVAAPTGYQLPSVTTAANLTGAVTASTAPAAATDT 567
QY 449 TNSDNLVNPTVNNKGTLPSTGGTGTTFYIIGAILVIG--AGIVLVARER 499
DB 568 -----TITNNKPFALPTGGQG-----LAGIIATVSGVIAFAIKR 604
```

RESULT 4  
A35259  
Type 1 fimbrial protein precursor - Actinomyces viscosus  
C:Species: Actinomyces viscosus  
C:Date: 10-Aug-1990 #sequence\_revision 10-Aug-1990 #text\_change 22-Oct-1999  
C:Accession: A35259  
R:Yeung, M.K.; Cisar, J.O.  
J. Bacteriol. 172, 2462-2468, 1990  
A:Title: Sequence homology between the subunits of two immunologically and functionally  
A:Reference number: A35259; PMID:90236904; PMID:1970561  
A:Accession: A35259  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-533 <YEU>  
A:Cross-references: GB:M32067; NID:g141849; PIDN:AAA62572.1; PID:g141850  
C:Keywords: transmembrane protein

Query Match 10.0%; Score 258; DB 2; Length 533;  
Best Local Similarity 26.0%; Pred. No. 9.1e-07;  
Matches 146; Conservative 67; Mismatches 223; Indels 126; Gaps 30;

```
QY 13 SLAFGMVSPVTPPIAFAAETGTITVODTKGATYKAYKVFDAAIDNANVSDSNKDGASYL 72
DB 17 TLAAGALVAP-TGAAPADPNNGST-DPDAATLITLVHKEQTDNTNGVKEGTGNE-D----- 69
QY 73 IPOGKEAYKASTDFNSLFT-----TTTNGGRTVV-TKDDTASANEI-ATWAKSISANT 124
DB 70 -PQ---AECKPVSDFEFTITKLVNVDLITTYDQWKLADLKGDDVVKAGALKSITVQKITGA 125
QY 125 TPVSTVTESNNDGTEVINVSQYGYVSVSTVNVGAVI-----MVT--SVTP-----N 169
DB 126 NGLASFTDAQ---TEV-----GAYLVSETRTPDKVIPAEDEFVVTLPMTNPQDTAKMNYN 176
QY 170 ATIEHKNTDWTGWDGGGKTV-DQKTSYSGVDTVKYTIYK-NAVNVHGTQKVVYQYVVKDTM 227
DB 177 VHVYPKNTLS-----GVDKQVTDKPAQSGGRDITITITSIPKVDYFGGARIKRVEVDRL 232
QY 228 -----PSASVVDLNEGSYEVTIDGSGNITLTQSGSEKATGKYNLLENNTFTIT 277
DB 233 DKRIKKEALTVPVKIV---GQNEVTLAE-TTDYTLIT-----AEGK-----DHNWATIQ 277
QY 278 IPWATNTPTGNTQNGANDDFYKGINT---ITVYTGVLKSGAKPGSADLPENTNIATI 334
DB 278 LTEGRRKASEARYNGNETKLVTLNAKFDAAVNLEGLDLSNTAGLI PNDSP-NFTWDPN 336
QY 335 NPNTSNDDPG---QKVTVRVDGQITIKKI---DGSTKASLQCAIFVL-----KNATGQFLNF 384
DB 337 NPGTTDIPGTPPTPVLSKYKGVLTKTGTDLADKTKYNGAQFOVYECTKTASATLUD 396
QY 385 ND--TNNVEWGTANATEYTTGADGIIITGLK-----EGTVYLVKCAP 427
DB 397 SDPSTQTVDPITIGEGTFTTAGQGTVEINLVANDYNGAKKOOLDDEYVCLVETKAP 456
QY 428 LGYN-----LDNSQKVLGSGATDTTNSDNLVNPTVNNKGTLPSTGGTGT 477
DB 457 EGYNLQADPLPFRVLAKEAK-----KAATEVTVD-----IPKNAGFRLPLTGTANGVI 505
QY 478 IFYIIGAILVIGAGIVLVARER 499
DB 506 FLTIAGALLVAGGAVVAYANKR 527
```

RESULT 5  
S52348  
Hypothetical protein 2 - Lactobacillus leichmannii  
C:Species: Lactobacillus leichmannii  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Nov-1999  
C:Accession: S52348  
R:Schenk-Groeninger, R.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: S52347  
A:Accession: S52348  
A:Molecule type: DNA  
A:Residues: 1-507 <SCH>  
A:Cross-references: EMBL:X81869; NID:g666067; PIDN:CAA57459.1; PID:g666069  
A:Experimental source: DSM 20076

Query Match 9.7%; Score 251; DB 2; Length 507;  
Best Local Similarity 24.7%; Pred. No. 2e-06;  
Matches 139; Conservative 78; Mismatches 204; Indels 142; Gaps 27;

```
QY 1 MKKK-----MIQSLVLASLAFGMVSPVTPPIAFAAETGT---ITV----- 37
DB 1 MKSKLSLITGLIMILPLLLLSFATPKVS-----AADTNNNSVKVTLHKRVFDSQAQEA 53
QY 38 -QDTOK-----GATYKAYKVFDAAIDNANVSDSNKDGASYLIPOGKEAYKAST 85
DB 54 KQNTGEIWNDFGGTGLNGVTFKAVNVDHYLSLRKSGSDAQDAVTAIOADKSDNLPYSY 113
QY 86 DFNSLFTTTTNGGRTYVTK--DTASANEIATWAKSISANTTPVSTVTESNNDGTEVIN- 142
DB 114 AGSATAATETT-----ATSKGEDGIAAFADNLNLKSDGNYQTYLVFVETDSPTDVTQQAAP 167
QY 143 -VSQYGYVSVSTVNVGAVIMVTSVTPNATHEKRVTDATWGDGGGKTVDQK-----TY 194
```



```
QY 436 SQKVIL---GDGATDTTNSDNLVNPVTVENNKGTLPSTGGIGTTFYIIG-AILVIGAG 491
Db 464 ASRAITVEPCAGVTQQWIDN--VKQSVPG-----LPLTGANGMLTLTASGAALLMIAYG 516
QY 492 IVLVARRRLR 501
Db 517 SVLVARYRER 526

RESULT 8
AB1347
probable peptidoglycan bound protein (LPXTG motif) lmo2178 [imported] - Listeria monocyt
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB1347
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1347
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1612 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00256.1; PID:gl6411648; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2178

Query Match 8.5%; Score 219; DB 2; Length 1612;
Best Local Similarity 24.2%; Pred. No. 0.00049;
Matches 116; Conservative 47; Mismatches 170; Indels 146; Gaps 23;

QY 43 GATYKAYKVFDAID---NANVSDSNKDGASY--LIPOG-----KEAEYK--ASTD---- 86
Db 719 GGYKSYNSTTGNIDWIVSNAMAKYNDLIFDDAIPTLGLTYVEGSLQYRNVASTSEMMN 778
QY 87 ----FNSLFTTTNGRTVYTKDITASAN---EIAFWAKS---ISANTPVSIVTESNND 136
Db 779 LVIPLNSVGTVAKTGDKNYPKVDITGKHLHLEFANLENSRFVFKYKTP----- 828
QY 137 GTEVINVSQGYGVYSSTV---NNG-----AVIMVTSVTPNATHEKNTDATWG 182
Db 829 ----NENWYFYVYQNTAKVSDNGVGEKSYQAVASKLFNMTKTATIDP----- 875
QY 183 DGGKTVDOKTYSVGDTVKTYTYKNAVNYHTEKYQVVIKDTWPSASVDLNGSYEV 242
Db 876 -----SFDKNVNIITLN---NISADRPINNPTITDTMKT-----GTTGA 912
QY 243 TITDGGSNITLITQSEKATGKYNLLENNNFITITIPMAATNPTGNTONGANDDFFYKG 302
Db 913 QVVKSSFKNWNETGEDIDSKYDIYTDNNFTIQPK-----DYKA 953
QY 303 INTITVYITGV-LKSG-----AKPGSADLPENINATINPNTSNDDPGQKV-----T 348
Db 954 TAPIKYTYTISIMSLVSNNTATTASPDYGLP-----MTYKSRTTISPAFTIGSGGT 1008
QY 349 VRDQGITIKKIDGSTKA-SLQGAIFVLKNATGQFLNFNDNNVWGTEANATEYTTGADG 407
Db 1009 ATIGSLEITKVDKONSKKLTGAKFOLYLTLEGD-----KAGQEA-----ITDSG 1053
QY 408 IITITGLKEGTYVLEKKAPLGNLLDNSQKVLGDGATDTTNSDNLVNPVTVENNKGT 466
Db 1054 KIYMDGLQSGKYLVEATEAPTGYTISDEYK-----DGKEITVTAD-IATSVTIENTEQT.1106

RESULT 9
AE1335
probable peptidoglycan bound protein (LPXTG motif) lmo2085 [imported] - Listeria monocyt
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AE1717
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
```

```
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AE1335
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1335
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-562 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00163.1; PID:gl6411555; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2085

Query Match 8.4%; Score 217; DB 2; Length 562;
Best Local Similarity 23.2%; Pred. No. 0.00016;
Matches 131; Conservative 55; Mismatches 197; Indels 182; Gaps 27;

QY 32 TGTITVQDTQKATYKAYKVPDAEIDNANVSDSNKDGASYL-----IPQKEAEYKAST 85
Db 80 TGNIEVWQNGNGYNVPAPF-----SGNFIELNSDGI GPVYQDIRTIP-GSNLTWKFS- 130
QY 86 DFNLSFTTTTNGRTVYTKDITASAN---SANIA-----TWAKSISANTT----- 125
Db 131 -----HRGRTGV---DTADLLIGSPESQTEVSRVSNGETWG-SPEGYNTVTPAGOT 176
QY 126 -----PVSTVTESNNDGTEVINVSQGYGVYSSTVN-NGAVIMVTSVTPNATHEKNT 177
Db 177 ITRLTFNPDISTANGSLTSGNFLDDVQLY-----INVNGA-----KIG 213
QY 178 DATWGDGGGKTVDOKTYSVGDTVKTYI-----TYKNAVNYHTEKYQVVIKDTWPSASV 232
Db 214 DVVWYDFNGDGIQDSEEPAPGPKVDLLTKDGTFKESAT---TNNIGSYLFTDVLDP---- 266
QY 233 VDLNNGSYEV---TITDGGSNITLITQSEKA-----TGKYNLLEEN---NNFTITIPW 280
Db 267 -----GDQVRFSLPNNDPFIISKANQNDKSLNSKPDKTGIASVNVNPNLKSSEFDI---- 317
QY 281 AATNPTGNTONG-----ANDDPFYKGINITITVYITGV-LKSGAKPG 321
Db 318 -----DAGITNGKVEIQKLSGDKALSGAVAIKDNSQSEVAKITTNQNGTGTAEGLPPG 372
QY 322 SADLPENINATINPNTS-----NDDPGQKVTRD-----GOITIKKIDGSTKASLOG 369
Db 373 NYTATEVTPAPLGQYQKNTTPFKFTITYGDPNPKLTFQNAEKTGSITIFKQDEANKKGLAN 432
QY 370 AIFVLKNATGQFLNFNDNNVWGTEANATEVTTGADGIITITGLKEGTYVLEKKAIPUG 429
Db 433 AVFDVKSIDGTYLK-----KVTNKSQYALAEALNLOPQGYVITEVTPAPG 476
QY 430 YNLLDNSQKVLGDGATDTTN---SDNLLVNP-----TV-----ENNKGTLPSTG 472
Db 477 YEKSANERIVTIPFPNQKNTINITFSDNKMIVBLKPTPKTGVSTVVKVSGSTTKITALPQTG 536
QY 473 GIGTTFIYIIGAILVIGAGIVLVAR 497
Db 537 DSSSSSTIFTGLLIVVASGLFVYRR 561

RESULT 10
AF1717
probable peptidoglycan bound protein (LPXTG motif) lin282 [imported] - Listeria innocua
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF1717
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
```

Science 294, 849-852, 2001  
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A;Title: Comparative genomics of *Listeria* species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AF1717  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1806 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC97510.1; PID:g16414794; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin282

Query Match 8.4%; Score 216; DB 2; Length 1806;  
Best Local Similarity 23.8%; Pred. No. 0.00083;  
Matches 114; Conservative 50; Mismatches 168; Indels 148; Gaps 19;  
QY 43 GATYKAYKVFDAEI-----DNANVSDSNKDGASYLIPOGKEAEYKASTD--- 86  
DB 721 GKGAYNSTGNIDMTIVSANAMAKSYDNLIFDDTIPTGLTYVEGSLQYRNVESSTEMMS 780  
QY 87 -----FNSLFTTTNGRTYVTKDTASANEIATWAKSISANTTPVSTVTSNNNDGTEVI- 141  
DB 781 LQIPLNSTGTAKAGDKNYPTKVDI-TGNKI-----HLEFANLIDNTRVFI 824  
QY 142 -----NVSOYGYVYVSSVYVNGAVIMVTSVTPNATHEKNTDATWGDGGKTVDOKTYIS 195  
DB 825 KYSTKPDENYFYKIVTN-----IAEVTONGT-----DKHTYS 857  
QY 196 -----VGDVTKYITTYKNAVNYHGT-----EKYQYVVKDTPSASVVDLNE 237  
DB 858 YQAVASVFNAMTKTASIDSTYNNKVNMTVLNISKDRPINNPITTDITM-----KN 909  
QY 238 GSVEVITDGSNITLITQSEKATGKXNLEENNFTHIIPWAAATVPTGNTQNGAND 297  
DB 910 GTTGAQVVKSSFRVNETDGDIDSKYDITFTDKOFTIOFK----- 951  
QY 298 FFYKGINITVYTG-V-LKSG-----AKPGSADLPENTNIATINPNTSNDPQOKVT 348  
DB 952 -DYKATDPIKVTYSTVSLMGLSNTATTATSPDYGSLPMSYKSRNTNISPAFTIGSGGT 1010  
QY 349 VRDQGITIKKID--GSTKASLQGAIFVLKNATGQFLNFNTNNVWTEANATEYTTGAD 406  
DB 1011 ATIGSLEITKVDKDNK-KLTGAKFQLYTPEGD-----KAGQEA-----TTDSE 1054  
QY 407 GIIITGLKEGTYVYVVEKKAPLGNLNDNSQKVLGPDGATDTTNSDNLNVPTVENNKGT 466  
DB 1055 KGIYMDGLQSGKYLVEAAPTGYTISDEYK-----DGKEITITAD-VATSVTIENTEQT 1108

RESULT 11  
AC1458  
surface anchored protein (LPXTG motif) [imported] - *Listeria innocua* (strain Clip11262)  
C;Species: *Listeria innocua*  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AC1458  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A;Title: Comparative genomics of *Listeria* species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AC1458  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-681 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC95435.1; PID:g16412621; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin0202

Query Match 8.3%; Score 214.5; DB 2; Length 681;  
Best Local Similarity 22.0%; Pred. No. 0.00028;  
Matches 110; Conservative 56; Mismatches 194; Indels 141; Gaps 16;  
QY 17 GMVSPVTPPIAFAETGIIITVODTQKATY-KAYKVFDAE-----IDNANVSDSNK 66  
DB 12 GLAFLSVLIISTVPTQTTIVKAATSYGSEFLNTVELLDKDGTPKTDGFGYDNNMNVH---- 67  
QY 67 DGASYLIPQKEAEYKASTDF-----NSLFTTTNGRTYVTKDTASANEIATW 116  
DB 68 --YTWISIPNSTNVKAGDSMDPALPOLALATDLAENVKDSKQTVGTATVTKRATNOV--- 122  
QY 117 AKSISANTTPVSTVTSNNND-----GTEVINVSQYGYVYVSSVTVN-NG 158  
DB 123 -----TVFSDYVEKHSIDKGLDFWTFPNQKVTGNEKINLE---PPIENSTINVDV 172  
QY 159 AVIMVTSVTPNATHEKNTDATWGGGG-----KTVDOKTYISVGDIVKTYI 204  
DB 173 NVGEKTPVSPTEITFYK-----WVDANNPSLIHWVVRVNYAOKNIPNAVFTDIIGAKQTL 228  
QY 205 TVKNAVNVHGTKEKVQYVVIKDTMPSASVVDLNEGSYEVTITDGSNITLITQSEKATGK 264  
DB 229 NFDISKAFHGTYSTDRVFTAGTPISTNPFKTSDFSVTL-----GNLTDVQLS----- 278  
QY 265 YNLLSENNNFITIPWAAATNPNTGNTQNGANDFFYKINGINTITVYTVGLKSGAKPGSAD 324  
DB 279 -----YTTATDGGKSTQYDNTAKLSGDFVTKTSTWTPASGG---GGGADGTT- 325  
QY 325 LPENTNIATINPNTSNDPQOKVTVRVDGQITIKKIDGSTKASLOGAIFVLKNATGQFLNF 384  
DB 326 -----GSVLTQDQAKTKATLEGAEFKLVANGTITLQE 358  
QY 385 NDTNNVWTEANATEYTTGADGIIITGLKEGTYVYVVEKKAPLGNLNDNSQKVLG 444  
DB 359 N-----ITDASGLNVNLSLKATYQLIETKAPTGYKLDSTPVEFTIGEN 403  
QY 445 ATD-TTNSDNLNVPTVENNK 464  
DB 404 NKEITVTKENTLDTGAVELTK 424

RESULT 12  
B98047  
hypothetical protein spr1403 [imported] - *Streptococcus pneumoniae* (strain R6)  
C;Species: *Streptococcus pneumoniae*  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C;Accession: B98047  
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; Es  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lerkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.B.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A;Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: B98047  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2551 <KUR>  
A;Cross-references: GB:AE007317; PIDN:AAL00207.1; PID:g15459054; GSPDB:GN00174  
C;Genetics:  
A;Gene: spr1403

Query Match 8.2%; Score 212; DB 2; Length 2551;  
Best Local Similarity 23.3%; Pred. No. 0.0021;  
Matches 121; Conservative 57; Mismatches 194; Indels 148; Gaps 24;  
QY 34 TITVQDTQKATYKAYKVFDAEIDN-----ANVSDSNKDGASYLI 73  
DB 1891 TFTVKDGKNGKGRAPKIKVEDITSPSIRRDAAATPRNGIRVTVVDDVNDNGVY-- 1948  
QY 74 PQGKEAEYKASTDFNSL-----FTTTNGRTYVTKDTASANEIATWAKSISANT 124



Db 1949 DSGVDKVLNSKDIYNGIDGRDGSAPTITTKDNGDGTHTITVQNPDSBSTTVVKDGDGK 2008  
QY 125 TPVSTVTESNNDGTEVINVSQYGYVSVGVNNGAVIMVTSVTPNATTHEKN-----TDA 179  
Db 2009 TANIITTE-NPDGSHITITVNPDSGTSTKETVKNKG-----DGKTPKEVETDNDGDGHTTKV 2063  
QY 180 TWGDCG-----GKTVDQKTSYVGDVTKYITITYKANVYHGTKEKYQYVLKD--- 225  
Db 2064 TDGDNVNTNAILKDGKGAAATATTENPDG-SHTVTITNP---DGTKN--EFVYKNGRD 2117  
QY 226 ----TPMSASVVDLNEGSVEVITDGSNGNITLTQSEKATKYNLLLENNTTITPWA 281  
Db 2118 GVDGRTPASVRDNGSGSHITVITNPEG-VTTEITVRDGSFKVITDB----- 2165  
QY 282 ANTPTGTQNGANDDFFYKGINTTITVTYGVLSKSGAKPGSADLPENNTATINNTSND 341  
Db 2166 -----QNGTHKISVLNGDGTET---IHKDGSFVATVRDQDGTIVRVENGNG 2213  
QY 342 DPGQKVTVRDGO-----ITIKKIDGSTRASLOGAIFVLKNATGQFLNFD 386  
Db 2214 TVSE--TTVRDGSKPTAKVVDNGDGTHTITVNSDGIITTT-----TVRDGREGPLEVID 2266  
QY 387 TNNVEMGTENANETTTGADGIITITGLKEGTYVLVEKKAPLGYNLLNSQ----- 437  
Db 2267 NN-----DGSHTIKVTGADGKGTITIPDG-----KSPKA-NIVDNGDGTHTLTIVD 2312  
QY 438 -----KVILGDGATDTTNSDNLNPNVTV---ENKKGTEL 468  
Db 2313 SDGREYKSIKDG-KDGKDS-----VSPTVTVKNNNDGTHV 2347

RESULT 13

AE1717  
probable cell surface protein (LPXTG motif) [imported] - Listeria innocua (strain Clp11  
C:Species: Listeria innocua  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AE1717  
R:Glauber, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A:Title: Comparative Genomics of Listeria species.  
A:Reference number: AE1077; MUID:21537279; PMID:11679669  
A:Accession: AE1717  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1622 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CMC97509.1; PID:gl6414793; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin2281

Query Match 8.0%; Score 205; DB 2; Length 1622;  
Best Local Similarity 22.4%; Pred. No. 0.0028;  
Matches 134; Conservative 65; Mismatches 207; Indels 192; Gaps 27;  
QY 18 MAVSPVTPIAFA---AETGTTV-----QDTQK---GATYKAYKVFDAIDNANVSDGNK 66  
Db 991 VAVSPVFTVGTSGTATIGTIKITKVDDEDDTTKLEGAKFQLY-TLDGKSEQEIT-TNS 1048  
QY 67 DGASYL--IPOGK-----EAQYK-----ASTDFNSLFTTTNG---GRTY 101  
Db 1049 EGEILLDGIQSGYKLVETKAPEGYNISDEYKEGKEITVSSSGEELLTIKNAMKGNVI 1108  
QY 102 VTKDQTASANEI-----ATWAKSISANTYPTVSTTESN----- 135  
Db 1109 LTKDSASHEVLADAEFELQNAATGSKLEKLTDAAGNIEITDLPAGDYKLIETKAPTGY 1168  
QY 136 --DGEV-----INVSOYGYVSVSVNNGAVIM-----VTSVTPNATTHEKNVTDATWG 182  
Db 1169 QLDATPVNFTIDFNQSEAAKVSKTNTAKTGTVWLTKDSATNAELADTFELRNEDETLV 1228

QY 183 DGGGKTVDQKTSYVGD-----TVKYTIITYKNA-----VNVHGTE 216  
Db 1229 RENLVTDDNGEISVADLAPGDYKLIETKAPTGYQLDAPVNFPTIDFNQSEVEKVSKTNTA 1288  
QY 217 KYVQYVI--KDTMPSASVVDL-----NEGS---YEVTTIDGSGNITLTITQSEKATGKN 266  
Db 1289 KGTGVVLTKDSDATNAELADATFELRNESDTLVRENLTVDNGEISV---ADLAPGDYK 1344  
QY 267 LLEENNNFTITPWAATNPTGTQNGANDDFFYKGINTTITVTYGVLSKSGAKPGSADLP 326  
Db 1345 LIE-----TKAPAGYQLDAPVNF----- 1363  
QY 327 ENTNATINPNATNSDDPGQKV-TVRDGQITITKKIDGSTKASLOGAIFVLKNATGQFLNPN 385  
Db 1364 -----TIDFNQSEAAKVSKTNTAKTGTVWLTKDSATNAELADATFELRNESDTLVREN 1417  
QY 386 DTNNVEMGTENANETTTGADGIITITGLKEGTYVLVEKKAPLGYNL-----LDNSQ 437  
Db 1418 -----LVTDDNGEISVADLAPGDYKLIETKAPTGYQLDAPVNFPTIDFNQ 1462  
QY 438 KVILGDGATDTTNSDNLNPN-PTVENNK-GTELPSTGGIGTTFYIIGAILVIGAGIV 493  
Db 1463 SEAAKVTNKKKIGTIIVNFIDVDGNLNDKEVHTGNVGE--YSVKAKEIVGHTLV 1518

RESULT 14

E95053  
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain TIC  
C:Species: Streptococcus pneumoniae  
C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: E95053  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: E95053  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-893 <KUR>  
A:Cross-references: GB:AB005672; PIDN:AAK74622.1; PID:gl4971932; GSPDB:GN00164; TIGR:SP46  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: Sp0462

Query Match 7.8%; Score 200.5; DB 2; Length 893;  
Best Local Similarity 20.6%; Pred. No. 0.0022;  
Matches 107; Conservative 69; Mismatches 164; Indels 179; Gaps 24;  
QY 69 ASYLIPQKKEABYKASTDFNSLFTTTTNGGRTYVTKDTSANETATWAKSI-----SAN 123  
Db 463 AAFPVKPEKYSKKAAG--YAVIGDPINGGYIWL-----NWRESILAYPFNSN 508  
QY 124 TTPVSTVTESNNDGTEVINVSQYGYVSVSVNNGAVIMVTSVTPNATTHEKNVTDATWG 183  
Db 509 TAKIT---NHGDPTP-----WYNGNIAPDGDYDFTVVGING--DPGTDEATAT 553  
QY 184 GGGKTVDQKT---YSVGDPTVK-----YTIYKANVYHGTKEKYQYVLKIDTSPAS 231  
Db 554 SFMQSISSKPENYNVTDTTKILEQLNRYFHTIVTEKKSIENT-----ITDPMGE 604  
QY 232 VVDLNEGS-----YEVTTIDGS-----GNITTLTQSEK---AT 262  
Db 605 LIDLQGLDGRFPADYTLTANDSGRLNGQAVGPGQNDGGLLKNKAKVLYDTTEKIRVT 664  
QY 263 GKNLLENNFTITIPWAATNPTGTQNGANDD-----FYKGINTTITVTYGVLSKGA 318  
Db 665 GLYLGTDEKVTLTYNV-----RLNDEFVSNKFF----- 692  
QY 319 KPGSADLPENTNIATINP-----NTSNDDPGQKVTVRD-----GQITTKK 358



```
Db 693 -----DINGRTTLHPKVEQNTVDRPIPK--IRDVRKYPEITISKEKLGIDIEFIK 742
QY 359 IDGSTASLOGAIFVLKQATGQFENFNTNNVWEGTEANAT--EYTTGADGIITITGLKE 416
Db 743 VNKNKKPLRGAVFSLQK-----QHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFFKNLSD 796
QY 417 GTYLVVEKKAPLGVLNLDNSQ-----KVILGD-----GATDTTNSDNLNVNPT 459
Db 797 GKTRLFENSEPAGYKPVQNPPIVAFQIVNGEVRDVTISIVPDIPAGVEFTNDKHYITNEP 856
QY 460 VENNKGTSLPGTGIGITTFIIGAILVIGAGIVLVARR 498
Db 857 IPPKR--EYPRTGIGMLPFYLGICMM--GGVLLVTRK 891
RESULT 15
B86807
hypothetical protein yoiC [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: B86807
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: B86807
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1441 <STO>
A:CROSS-references: GB:AE005176; PID:gl2724450; PIDN:AAK05556.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yoiC
Query Match 7.8%; Score 200; DB 2; Length 1441;
Best Local Similarity 23.9%; Pred. No. 0.0045;
Matches 132; Conservative 61; Mismatches 199; Indels 160; Gaps 30;
QY 23 VTPIFAAEAGTIT-----VDQTK--GATYKAYKVPDAEIDNA 59
Db 799 VTPTALAGQNTTDTYSVVQYNTNTIKGISIVPVDSGTGAIANAKATAYEKYSDETVNA 858
QY 60 NVSDSNKDGASYLIPQGEAEYKAST--DFNSLFTTTNGGRTYVTKDPTASANEIATW- 116
Db 859 PTAP-----AGYVL--NDVNNKAVSVADANGVINFYDAIQTLTKD-STIDEGSTWT 909
QY 117 -AKSISANTTPVST-----VTESNNDGTEVINVSQYGYVSVSTVNNGAVIMVTSVTPN 169
Db 910 AADNFTGGTSGTGLVINDVTGTVDTPGVSSTVYTYTDP--TGASI---SSVAN 964
QY 170 ATTHEKN--TDATWGDGGKT--VDQKTY-----SVGDTVKYTYTKYKNAVNYHGTKE 217
Db 965 ITVNDSSNTNPTDGTGGGNTNPTDGNITVLPGQDKLTADDVTVGQDPLSPSDGS 1024
QY 218 VYQYVIKDTMPS-ASVVDLNEGSEYV---TITDGSNI-----TTLTQGEKATGKNLL 268
Db 1025 V-----TLPSDGGKVDKPDGSYNVPGGTWDPDGTIHLPGGTVINPGG----- 1067
QY 269 EENNFTITIPWAATNPTGNTQNGANDPFYKGINTI-----TVTYTGLVKSAGKES 322
Db 1068 -----SVTVP-----GPDGKT--GTDDDTTLNPNSPVPVPGDNGSVTLPGGGTASTPNGN 1114
QY 323 ADLPENTNIATINPNTSNDPQKQVTVRDGGITIKIDGST-----KASLOGAIFVLKN 376
Db 1115 IITLPGGT--VVPDGTIHLPGGDIVNPDGTITLPGDGKGTGDDGKIKPNGLIPGDN 1171
QY 377 AT-----GQFLNFDNNVWEGT--EANATEYTTGA-----DGIITITGLKEGTYLVE 423
Db 1172 GSVTLPGGGTVTTPGGGTINVPGGSVDPDGTVHLPGGDIVNPDGTITLPG----- 1221
QY 424 KKAPLGYNLLDNSQKVLGDGATDTTNSDNLVN-PTVENNKGTP-ELPSTGGIGTITFYI 481
```

```
Db 1222 -----QDGKGTGTGDDGKVKPNGPSISNPDGSITLPGGGTVTTP----- 1259
QY 482 IGAILVIGAGIV 493
Db 1260 GGTINVPGGSVV 1271
RESULT 16
AH1396
peptidoglycan anchored protein (LPXTG motif) [imported] - Listeria monocytogenes (strain
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH1396
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1396
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1530 <GLA>
A:CROSS-references: GB:NC_003210; PIDN:CAD00654.1; PID:gl6412064; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2576
Query Match 7.7%; Score 199; DB 2; Length 1530;
Best Local Similarity 21.7%; Pred. No. 0.0055;
Matches 141; Conservative 74; Mismatches 192; Indels 244; Gaps 33;
QY 18 MAVSPVPIAFAAETGTTIVQDTQKATYKAYK-----VFDAEI--DNANVS 62
Db 675 MLITFKKLIHPVE---VTYKTPVGITKPLPKYKAVISDGEVLADYEAEVIDNAN-- 729
QY 63 DSNKDGASYLIPQGE-----AEYKASTDFNSLFTTTNGGRTYVTKDPTAS---- 109
Db 730 -----KYVNSGEGVGNIDWEIYANGSGSTVSNATVTDLTGTGQ-----KLTSSIKV 778
QY 110 -ANEIATWAKSISANTTPVS-----TVTESNNDGTEV-----INVSYGYVSVST 154
Db 779 YKSQTSVTKMLQESNMPISPGEYDLKTGVDEESLVEYFQVKFNEINOSYVIKYQTAIT 838
QY 155 -----VNGGAVIMVTSVTPNATHEKNTDA--TWGDCGG-----KTVDQ-- 191
Db 839 LITSDTETTAQIGNSVTFGDNITKGETEKNIEVKITTGDTGTGETGKIILNKVDKAD 898
QY 192 -----KTSYVGDTVKYTYTKYKNAVNYHGTKEVY-QYVIKD-----TMPASV- 232
Db 899 PSIPLEGATFDLYANDEKVDQTQTDKNGV-IEFDLLVYGDYTLKEVSAPEGVTLPTASTE 957
QY 233 ---VDLNEGSEYVITD-----GSGNITLTLQGEKATG-----KYNL----- 267
Db 958 NIQVKLEODEKVVQVNNKEMPIKETGEVHLVK--TDKATGATLAGAEFSLDKSGAELQN 1015
QY 268 ---LEENNFTI-----TIPWAATNPTG-----NTQNG-----ANDDFP 299
Db 1016 GLTTDENGELTIHNLDLGSYLYKETKAPGKYLSEKTWEFSVESGQVDAIEIOAENEKDL 1075
QY 300 YGINTITVTVYGLVKSAGK-----PG-----SADL 325
Db 1076 GEAVLTQVDSKSETNAKLSGAKFNLLNDSGEVIQTNLVSDENGEIRVQNLEFGDYAFQETA 1135
QY 326 PENTNIATINPT-----SNDPQKQVTVRDGGITIKIDGSTKASLQ 368
Db 1136 PTNYDIAT---NTPPTIVAGQTSATWTAENKKTG-KPDVDTGEVILVQDSATGETLE 1191
QY 369 GAIFVLKNATGQFLNFDNNVWEGTEANATEYTTGADGIITITGLKEGTYLVEKKAPL 428
Db 1192 GAVFDLMTADGAIV-----ASNLTTDANGEITVTNLPAGKYSFKETKAPE 1236
```

QY 429 GYNLLDNSOKVILGDGATDT---TNSDNL---LVNPTVNNKGTFLPSTGGI 474  
Db 1237 GYEL-----ATDWBEETAPNQPEKIIITAENTKLAIPIDAGSV 1275

RESULT 17  
AII094  
probable peptidoglycan bound protein (LPXTG motif) lmo0160 [imported] - Listeria monocytogenes  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AII094  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, P.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AII094; MUID:21537279; PMID:11679669  
A:Accession: AII094  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-571 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC98375.1; PID:gl6409519; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0160

Query Match 7.5%; Score 192.5; DB 2; Length 571;  
Best Local Similarity 21.4%; Pred. No. 0.0034; Mismatches 191; Indels 191; Gaps 26;  
Matches 122; Conservative 65; Mismatches 191; Indels 191; Gaps 26;

QY 46 YKAYKV-FDAEIDNA-NVSDSNKDGASYLIPQCKEAYKASTDFNSLFTTTNGGRT--- 100  
Db 64 YDSIKVHYTWEPNSTNVKEG-DTWEFVLP---PELKIIVTDLD--FSLKDHGDNVTGCH 115

QY 101 YTKDKDTASANEIATWAKSISANTTTPVSTVTSNNDGTVEINVQGYYY-----VSS 153  
Db 116 VIAKSKTGQV-VITTFDFEKNSS-----NISGYLDFWNWDRSLVEG 156

QY 154 TVNNGAVIMVTSVTPNATHEKNTDATMGDGGKTV--DQKTVSVG-----DTVKYTI 204  
Db 157 NENPVVEFPNGTETIDV-----GVGGKQIDPDESLYKGWADAENPELIQWV 207

QY 205 TYKQNAVNYHGTBKVOYVIKDTMPSASVVDLN-----EGSYEV--TITDG-----SGNITT 253  
Db 208 ----RVNY-SKQIQNAVVEDFIGPVQVDFNSIKAFHGFEPDDNFTFGAEPSSAITQ 262

QY 254 LTQSEKATGKNLLBENNFTITIPWAAINTPTGNTQNGAN-----DDFPYKGINNTIT 307  
Db 263 TTDGFKVDLG--NLTD-----SVKISYTTSTDNGASPNYTKGQLTGDNFIKQIEVAT 315

QY 308 VITYTVGLKSGKPGSADLPENTNIATNP-----DDPGQKV-----NTS 339  
Db 316 PTSGGGGGEGTGTGSVELTKTDSSQKNLEGAEBFLVNGAGATVGTGLKTWNADGKLATS 375

QY 340 N-----DDPGQKV-----TVRDQGITIKKIDGSTKA 365  
Db 376 NLKYDTYQLLETKAPOGYVLDASPVKFTIDDTQSLFVSKENTAIKGSVSLKVDRTKS 435

QY 366 SLOGAIFVLKNATGQFLNFDNTNNVEWGEANATEYTTGADGIIITGLKEGYIYLVKK 425  
Db 436 LLADADEFELQDKDGNLT-----TNLKTQKGLTVADLLPGEYQVETK 480

QY 426 APLGYNLLNSOKVILGDGATDTTNSDNLVNPTV-----ENNKGT 467  
Db 481 APTGYILDTTLPFKISTEALNVTVTKENTKPEIPKVPVPPKPKPKKIISSESKQTA 540

QY 468 LPSTGGIGTTFIIGAILVIGAGIVLVA 496  
Db 541 LPKTG-----DSPLVNGWGLLVA 559

RESULT 18  
AB1283  
peptidoglycan linked protein (LPXTG) [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AB1283  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, P.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AB1283  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1711 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC99744.1; PID:gl6411102; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo1666

Query Match 7.4%; Score 191; DB 2; Length 1711;  
Best Local Similarity 21.7%; Pred. No. 0.017; Mismatches 127; Conservative 73; Mismatches 184; Indels 200; Gaps 28;

QY 17 GMAVSPVTPIAFAAETGTTTODTKGATYKAYKVDFDAEIDNANVSDSNKDGASYLIPQ 76  
Db 1226 GNKADPVII-----TIVVADTEK-----PIITADTTTYAKGTTKTVQAQFL----- 1266

QY 77 KEAYKASTDFNSLFTTTNGGRTYVTKDTSANEIATWAKSISANTTTPVSTVTSNND 136  
Db 1267 -----TDIHA-----TTNDGSTIISNFPDTPVLAQEGY-----TVVNAKDESNNE 1307

QY 137 GTEVINVSQGYGYVSVSTV--NNGAVI-MVTSVTPNATHEKNTD-----ATWGDGGK 187  
Db 1308 ADPVT-----VTITVVDTKGPIIALNALAIYERTINKNEADFLADIEATTDDGSTI 1358

QY 188 TVD-----QKTVSGDVTKYTITYKNVNYH-----G 214  
Db 1359 TTFNSKDLDTVGTIVYVILNAEDASGNKATPVKTIKVEDTIPPIITADQSITYERGITK 1418

QY 215 TEKVYQYVIK-----DTMPSA---SVVDLNE-GSYEVTI--TDGSGNITLTQSEKATG- 263  
Db 1419 TEQAFTYDIKAATSDNSPISDFSKIDLTKTGNVEVLLRATDQSGN-----KALPL 1469

QY 264 KYNLLBENNFTITIPWAAINTPTGNTQNGANDDPYKGINNTITVITYTVGLKSGAKPGSA 323  
Db 1470 KINVLVQD-----TIAPVKTTSREITA 1492

QY 324 D-----LPENTNIATNPNTSNDPQKVTV-----RDQGITIKKIDGSTKASLQ 369  
Db 1493 ERGTPTWTEOQLAKIGANT---DQSGKITTDYNPALVNTSGDYLVHLYAVDAAGNQAI 1549

QY 370 AIFV-LKNATGQFLNFDNTNNVEWGEANATEY-----TTGADGIIITIG-----LKEGT 418  
Db 1550 EITIHVKDTPAPIKADKKISYPVGVKTVTFEQLDIIHATTDDGSKITTFDFPNMLKTPG 1609

QY 419 YVLV-----EKAPLGYNLLNSOKVILGDGATDTTNSDNLV 456  
Db 1610 KYTIHLNVAADVADGNKAKTIDVSLTVEEKVTPPKPTPDDGGNNTGNSNGTSNTD-ITV 1668

QY 457 NPTVNNKGT--ELPSTGGIGTTFIIGAILVIGAGIVLVARR 498  
Db 1669 NPTKONTATNESIPALGDTKSTIPVIGMFL-LATSLVLIRRK 1711

RESULT 19  
S41525  
major ring-forming surface protein precursor - Helicobacter mustelae  
C:Species: Helicobacter mustelae



Query Match		7.2%; Score 186; DB 2; Length 3705;
Best Local Similarity		23.5%; Pred. No. 0.088;
Matches 112; Conservative		70; Mismatches 213; Indels 82; Gaps 21;
QY	33	GTITVQDTQKGATYK---AYKVEDAEID--NANVSDSNKOGASYLIPQGEAEYKASTD 86
DB	473	GIINTSATANGITPAGTEGGHILDTLILNLGTGIALSNVAGVNLTL-----SNT 523
QY	87	FNSLFTTTNG--GRITYVTKDTASANEIATWAKSISANTTPVSTVTESSNNNGTEVINVS 144
DB	524	LNTLANGTALNSLTGLTLDVSLNGRNTINIEGAGIGIAATNTELTNTEFAEALD----INVN 579
QY	145	QVGYVYVSTVANGAIVMTVSTVPNATIEHKNTDAT---WGDGGKTVQKYSVGDTVK 201
DB	580	GAG---IGIQTGGGVNLSAS---NLIINVANTLGTALQITDG-----IDNTT-TIGNEIQ 628
QY	202	YTIYKNAVNYHGTEKV---YQVVIKDTMPSASVVD---LNEGSEYBVTITDGSNNITLT 255
DB	629	LNAENATALNPLGSSKTLNNGTIKGSVIFAGVADHIIINNGTLDGLTITGAGNDTLVL 688
QY	256	QGSEKATGKYNLLBENNFTIIPWAATNPTGNTQNGANDDPFYKGINTIITVYTGVLK 315
DB	689	DSSQSNQVNLGDGNSVTIQNGATVSSIIIGN---GNDTFTNGM-SVGSTYLGSLD 743
QY	316	SGAKPGSADLPENTNIATINPTNSDDPGQKTVR-----DGOITIKKID--GSTKA 365
DB	744	AG-----TGLNTLNFNASTDELAASLOQFTNINLVDSHITLVSDDNIGSGMV 792
QY	366	SLQGAIFVLKNATQQLNFNDTNVWEGTEANATEYTTGADGIIITITGLKEGYVILVEKK 425
DB	793	NIDSSSELLFGST-----FDGILHATLGAGTSGAIVNNSANVSLEQASMPAGTWQVNOGG 847
QY	426	APLGVNLLDNSOKVILGD---GATDTTNSDNLVNPTVENNKGTELPSFGGIGTTF 479
DB	848	A-----LTSNSNLGSAKIGLDGTLNLDNLALFNHVLGTNGT-LNVAKNLATTA 897
RESULT 22		
S57721		
cspB protein - Clostridium acetobutylicum		
A:Species: Clostridium acetobutylicum		
C:Date: 19-Oct-1995 #sequence_revision 17-Nov-1995 #text_change 15-Oct-1999		
C:Accession: S57721		
R:Sanchez-Beato, A.; Garcia, J.		
submitted to the EMBL Data Library, July 1995		
A:Description: Molecular characterization of a family of choline-binding proteins of Clo		
A:Reference number: S57714		
A:Accession: S57721		
A:Status: preliminary		
A:Molecule type: DNA		
A:Residues: 1-583 <SAN>		
A:Cross-references: EMBL:Z50009; NID:g895758; PIDN:CAA90304.1; PID:g895759		
C:Superfamily: cpl repeat homology		
F:504-523/Domain: cpl repeat homology <CP2>		
F:544-563/Domain: cpl repeat homology <CP3>		
Query Match		7.2%; Score 185.5; DB 2; Length 583;
Best Local Similarity		21.6%; Pred. No. 0.0082;
Matches 132; Conservative		83; Mismatches 215; Indels 181; Gaps 28;
QY	1	MKKKMTQ---SLLVASLAFGMVSPVTPPIAFAAETGTTVQDTQKGATYK--AYK-----VF 52
DB	1	MIKRMTKATSLVAA-----AAIISIVP-----AHAADYTKIDSQEGTIYDAVAYKDGKFV 52
QY	53	DAEIDNANVSDSNKOGASYLIPQK-----EAQYKASTDF-----NSLFTTTNG 97
DB	53	DGEV-----NDKDEAAAYLADGKYNLSIDSGADTDIYGSKYLEVQDGYFDVLDNG 105
QY	98	GRITYVTKDTASANEIATWAKSISANTTP--VSTVTESSND--GTEV-----INVSOYGY 149
DB	106	SVTDESIKENAEDDAASALRKNLKXNDKRYASSERETIQLAGAIAGNKFSAPYKFE 165
QY	150	YVSSVYVNGAV-----IMVT-----SVTPNATIEHKNTDATW-GD 183

DB	166	YTKDNATNGSVAKLVNYTDAQGNYYIDADYNLGSIKVTTTASSTTSKTVTIANTDSDYDAD 225
QY	184	GCGKTVQDKYTSVGTQVYKITYKNAVNYHGTEKVYQVYVVKDTMPSASVVDLN-----E 237
DB	226	GTSDATGKKSASVDTSGKVLIGQDANYIYRTAK-----VTVNAANGATITKINGMDVTVD 281
QY	238	GSYEYVITDGSNITITLQGSSEKATGKYNLLBENNFTIIPWAATN-----TP 286
DB	282	CGHTFTVSNQ-----TLGSSASTVGFNVIQK-----ISKAQASGVNDGAKYAKSVTTY 329
QY	287	TGNTONGANDDPFY-----KGINTIIVYITGVLKSG-----AKPGSADL 325
DB	330	IYVSDGKNEAFAYDNYTVANGKIVGYTANGTNVNTATGTLSSKNGYVYVLDGLDEASEDV 389
QY	326	PENTNIATINPT-----SNDDPGQKVTVRDG---QITIKKIDGSTKASL 367
DB	390	KENGTKSAVDTDVCGKWRDLGAGIYIKWDNDEDTWKYKVDGSGFDQMSVYNQDNIVANWSK 449
QY	368	QGAIFVLKNATQQLNFNDTNVWEG---TEANATEYTTG-----A 405
DB	450	EDVYSVIGGKGTSTPTDTPVNVKGWKTDAGWTFYNTDPTQVKQWVNDGQVWYIYKA 509
QY	406	DGIIITITGLKECT--YVLVEKKAPLYNLLDNSQKVL--GDGATDT-----TN 450
DB	510	DGTMAATGIDGSTWYLYLQSSGAMKTGWLNDNGTWYLYLQSSGAMSTGWLNDNGTWYFLNS 569
QY	451	SDNLLVNPVTE 461
DB	570	SGMLANTVTD 580
RESULT 23		
A48995		
paracrySTALLINE surface layer protein ResA - Caulobacter crescentus		
C:Species: Caulobacter crescentus		
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000		
C:Accession: A48995		
R:Gilchrist, A.; Fisher, J.A.; Smit, J.		
Can. J. Microbiol. 38, 193-202, 1992		
A:Title: Nucleotide sequence analysis of the gene encoding the Caulobacter crescentus pa		
A:Reference number: A48995; MUID:93007489; PMID:1393820		
A:Accession: A48995		
A:Status: preliminary		
A:Molecule type: nucleic acid		
A:Residues: 1-1026 <GIL>		
A:Cross-references: GB:AF062345; GB:M22663; GB:M84760; NID:g6064104; PIDN:AAC38665.2; PI		
A:Experimental source: CB15A, ATCC 19089		
A>Note: sequence inconsistent with the nucleotide translation		
A>Note: sequence extracted from NCBI backbone (NCBI:116173, NCBIP:116174)		
Query Match		7.2%; Score 185.5; DB 2; Length 1026;
Best Local Similarity		22.1%; Pred. No. 0.017;
Matches 131; Conservative		66; Mismatches 240; Indels 157; Gaps 26;
QY	6	IQSLVASLAFGMVSPVTPPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNANVSDSN 65
DB	299	VQAAAVTALPTGVTISGIE--TMNVTSGAAITLNTSSGVT-----GLTALNTN 344
QY	66	KDGASYLIPQGEAEYKASTDFNSLFTTTNG-----RTVYTK-KDTASANEIATWAK 118
DB	345	TSGAAQTVTAGAGQNLTLATTAQAANNVAVDGGANNVASTGVTGTTTGGANSASGTV 404
QY	119	SIS-ANTTIPVST-----VTESNDCGTEVINVSQYV 148
DB	405	SVSVANSSTTTTCAVATGGTATVTAQTAGNAVNTLTQADVTVTGNSSTTAVTQTAA 464
QY	149	YVSSVTN---NGAIVMTVSTPNATIEHKNTDATWBGCGGKTVQKYSVGTGTVKTYIT 205
DB	465	ATAGATVAGRVNGAVTITDSAAASATTAGKIATVTLGSFGAATIDSSALT----- 514
QY	206	YKNAVNYHGTEK---YQVYVVKDTMPSASVVDLNGSGSEYV---ITDGG-----NITTLTQ 256

Db 515 ---TVNLSTGCTSLGIGRGALTAT-PTANTLTNLVNGLTITGTAITDSEAAADGGFTTINI 570  
Qy 257 GSEKATGYNLLBENNFTTITIPWAATNTPTGNTQANGDDPFYKGINITVT----- 309  
Db 571 AGSTASSTIASLAADATTLNISGDARVITTSHTA-----AALTGIIIVTSVGATL 621  
Qy 310 ----YTCVLKSGAKPGSADLPENTNIA-----TINPNTSNDPQGVTVVRDG-QIIT 356  
Db 622 GAELATGLVFTGGAGADSIILGATTKRAIVMGAGDDTVTVSSATLGAGGSVNGDGDVVLV 681  
Qy 357 KKIDGSTKASLOGAIFVLKKNATGQF-----LNFDNTNNVEMWGTEANATEYTT 403  
Db 682 ANVNGSS-----FSADPAFGGFETLRVAGAAQAQSHNANGFTALQLCATAGATTFTN 733  
Qy 404 GADGI-ITITGLKEGTYYLVEKKAPIG-----YNL-LDNSQKVILGD-----GAT 446  
Db 734 VAVNGLTVLAAPTGTITVTTLANA-TGTSDFVNLTLSSSAALAAGTVALAGVETVNIAT 792  
Qy 447 DTTNS---DNLLVNPT-----VENNKGTLPSTGGIGTTFIYIIGAILVIGAG 491  
Db 793 DTNTTAHVDTLTLOATSAKSIVVTGNAGLNTNTGNTAVTSF---DASAVTGTG 843  
RESULT 24  
C87374  
S-layer protein Reaa [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: C87374  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1073 <STO>  
A:Cross-references: GB:AB005673; NID:g13422297; PIDN:AAK22991.1; GSPDB:GN00148  
A:Genetics: C81007  
Query Match 7.2%; Score 185.5; DB 2; Length 1073;  
Best Local Similarity 22.1%; Pred. No. 0.018;  
Matches 131; Conservative 66; Mismatches 240; Indels 157; Gaps 26;  
Qy 6 IQSLLVASLAFGMVSPVTPPIAFAETGTTIVQDTQKATYKAYKVPDAEIDNANVSDSN 65  
Db 346 VQAAAVTALPTGVTISGIE--TMNVTSAAITLNTSSGVT-----GLTALNTN 391  
Qy 66 KDGASVLPQKKEAEYKASTDFNSLFTTTTNGG-----RTYVTK-KDTASANEIATWAK 118  
Db 392 TSGAAQVTTAGAQNLTAATAQAANNVAVDGGANVTVASTGVTSGITTVGANSAAAGTV 451  
Qy 119 SIS-ANTTPVST-----VTENNNDGTEVINVSQY 148  
Db 452 SVSVANSSTTTTGAAVAGTGTAVTVAQTAGNAVNTLLQADVTVTGNSSTTAVTQTAA 511  
Qy 149 YVSVSTVN---NGAIVMTSVTPNATIEKNTDATWGGGKTKVDOKTYSVGDIVKYYIT 205  
Db 512 ATAGATVAGRVNGAVTITDSAAASATAGKIATVTLGSGAATIDSSALT----- 561  
Qy 206 YKNAVNYHGTGK---YVQVVKDTMPSASVVDLNEGSYEVT---ITDGGG---NITTLTQ 256  
Db 562 ----TVNLSTGTSIGIGRGALTAT-PTANTLTNLVNGLTITGTAITDSEAAADGGFTTINI 617  
Qy 257 GSEKATGYNLLBENNFTTITIPWAATNTPTGNTQANGDDPFYKGINITVT----- 309  
Db 618 AGSTASSTIASLAADATTLNISGDARVITTSHTA-----AALTGIIIVTSVGATL 668  
Qy 310 ----YTCVLKSGAKPGSADLPENTNIA-----TINPNTSNDPQGVTVVRDG-QIIT 356

Db 669 GAELATGLVFTGGAGADSIILGATTKRAIVMGAGDDTVTVSSATLGAGGSVNGDGDVVLV 728  
Qy 357 KKIDGSTKASLOGAIFVLKKNATGQF-----LNFDNTNNVEMWGTEANATEYTT 403  
Db 729 ANVNGSS-----FSADPAFGGFETLRVAGAAQAQSHNANGFTALQLCATAGATTFTN 780  
Qy 404 GADGI-ITITGLKEGTYYLVEKKAPLG-----YNL-LDNSQKVILGD-----GAT 446  
Db 781 VAVNGLTVLAAPTGTITVTTLANA-TGTSDFVNLTLSSSAALAAGTVALAGVETVNIAT 839  
Qy 447 DTTNS---DNLLVNPT-----VENNKGTLPSTGGIGTTFIYIIGAILVIGAG 491  
Db 840 DTNTTAHVDTLTLOATSAKSIVVTGNAGLNTNTGNTAVTSF---DASAVTGTG 890  
RESULT 25  
A83412  
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83412  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miziochuchi, S.D.; Warrenner, P.; Hickey, M.J.; Br.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: A83412  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2468 <STO>  
A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN0013  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA1874  
Query Match 7.2%; Score 184.5; DB 2; Length 2468;  
Best Local Similarity 21.0%; Pred. No. 0.062;  
Matches 116; Conservative 69; Mismatches 209; Indels 159; Gaps 24;  
Qy 21 SPVTPIAFAAETGTI---TVQDTQKATYKAYKVFDA-----EIDN-ANVSDSNKDG 68  
Db 1044 TFSFPLA---DGTVVNATATDPAGNTGGGSTTVDAIAPATPTVNLNSGSSLSGTAEFG 1099  
Qy 69 ASYLIPOGKEAEYKASTDFNSLFTTTTNGR---TYVTKDTASANEIATWAKSISANTTP 126  
Db 1100 STVILTGD-----NGNPIAEVTADGSGNWTYTPPIANGTVVNVVVAQDASGNSP 1150  
Qy 127 VSTVT-ESNNDGTEVINVSQYGYVYSVTNNGAVIMTSTVTPNATIEKNTDATWGDGG 185  
Db 1151 PATVTVDSAPPAPVNPIS-----NGVVISGTA-----EAGATVTLTDAG 1190  
Qy 186 GKTVDQKT-----YSVGDTV-KYTIITYKNAVNYHGTGKVKYVYVVKDTM-PSASVVDL 235  
Db 1191 GNPIGVTTADGSGNWSFTPGTPLANGTVIVATATDPTGNTGPOAATTVDAVAPPVIDP 1250  
Qy 236 NEGS-----YEVITDGSNIT-----TLTQSEKATGYNLLENNFT 275  
Db 1251 SNGTTSIGTAEAGAKVILTNGNPNIGETTTADGSGNWSFTPGTPLANG----- 1298  
Qy 276 ITIPWAATNTPTGNT---QNGANDDFYKGINITTVTVTVGLKSGAKPGS----- 322  
Db 1299 -TVNVAQDPAGNTGPGQSTTVDAVAPNTPVNPSNGNLLNGTAEPGSIIVTLTDGNGP 1357  
Qy 323 -----ADLPENTNI-ATINPNTSNDPQGVTVVRDGQITIKKIDGSTK 364  
Db 1358 IGQTTADGSGNWSFTPGSQLPNGTVNVNVTASDAAGNTSLPAITTVDSLSLPSIPQVDPNSG 1417  
Qy 365 ASLQGAIFVLKKNATGQFLNFDNTNNVEMWGTEANAT---EYVTCADGIIITIT---GLKEGT 418  
Db 1418 SVISGTA-----DAGNTIIITDGNPNIGQVTTADGSGNWSFTPGIPLPDGT 1463  
Qy 419 YVLVEKKAPLGYNLLONSQKVILGDGATDTTNSDNLVNPVTENNKGTELPSTGGIGTTI 478

Db 1464 VNVVARS---SNVDSAPAVITVDGVAPAA-----PVIDPSNGTISGTAERAGATV 1512  
QY 479 FYIIGAILVIGAG 491  
Db 1513 -----ILTDGGG 1519

RESULT 26  
T28679  
fibrinogen-binding protein homolog - Staphylococcus aureus  
C:Species: Staphylococcus aureus  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C:Accession: T28679  
R:Josefsson, E.; McCrea, K.; Ni Eidhin, D.; O'Connell, D.; Cox, J.; Hook, M.; Foster, T.  
Microbiology 144, 3387-3395, 1998  
A:Title: Three new members of the serine-aspartate repeat protein multigene family of St  
A:Reference number: Z20510; MUID:99098700; PMID:9884231  
A:Accession: T28679  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1315 <JOS>  
A:Cross-references: EMBL:AJ005646; NID:e1318791; PID:e1318792; PIDN:CAA06651.1  
C:Genetics:  
A:Gene: sard

Query Match 7.0%; Score 181.5; DB 2; Length 1315;  
Best Local Similarity 22.0%; Pred. No. 0.039;  
Matches 116; Conservative 72; Mismatches 191; Indels 149; Gaps 29;

QY 34 TITVQDTQKATYKAYKV--FPAEDINAVNSDKGAS---YLIPQKEAE--YKAST- 85  
Db 427 TIYNPSENSLNAKLKVQAYHSSYPN-NIGQINKDVTDIKIYQVPKGYTLNKGYDVNTK 485

QY 86 -----DENSLETTTNGGRTYV-----TKDKTASANEIATWA--- 117  
Db 486 ELTDTVNQYLKITYGDNNSAVIDFGNADSAYVMWNTKFQYTNSESPLVQMATLSSTG 545

QY 118 -KSISANTTTPSVTBSNNDGTEVINVSQYGYVYSSTVNNGAVIN---VTSVTPNATI 172  
Db 546 NKSVSFGNALGFTNNSGCGAQEVYKI---GNYWEDTNKNGVQELGEGKGVNVT--VTV 600

QY 173 HEKNTDATWGGGKTVDOKTYSV-----GDTVKIITIYKN----- 208  
Db 601 FDNNTNWKVGE--AVTKEDGSLIPNLPGD---YRVEFSNLPKGYEVTTPSKQGNNEELD 655

QY 209 -----AVNYHGTKE-----VYQ-----YVQKTMPSASVDLNE---GSYEVTITD 246  
Db 656 SNGLSVITVNGKDNLSADLGHYKPKYNLGDYVWEDTNKN-GIQDQEKGISGVTVTLKD 714

QY 247 GSGNITTLTQSGSEKATGKYNLLE-ENNFTIITIPWAAATNPTGNTONGANDDFFYKGINT 305  
Db 715 ENGNV--LKTVTTDADGKYKFTDLONGYKVEFTTPEGYTPTVT--SGSDIEKDSNGLTT 771

QY 306 IIVTVTGVLKSKAGPSADLPENTNIATNPNTSNDPQKV---TVRDGQITIKIDGS 362  
Db 772 -----TGVI--NGA-----DNMTLDSGFYKTPKYNLNGYWEDTNKDGK-----QDS 811

QY 363 TKASLOGAIFVLKNATGQFLNFDNTNVEWTEANATEYTTGADGIITITGLKEGYTLV 422  
Db 812 TEKGISGVTVTLKNEVGEVLQTTTKDK-----DGKYQFTGLENGT-YKV 854

QY 423 EKKAPLGYNLLDNSQKVLGDGATDTTNSDNLNVPVNNKGTPLPS 470  
Db 855 EFETPSGYT-----PTQVSGTGDEGIDNSNGTSTTGVIKDKNDITDS 896

RESULT 27  
S76109  
hypothetical protein - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S76109  
R:kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.  
A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76109  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3029 <KAN>  
A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10087.1; PID:d1010731  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 7.0%; Score 181; DB 2; Length 3029;  
Best Local Similarity 20.5%; Pred. No. 0.13;  
Matches 117; Conservative 63; Mismatches 199; Indels 192; Gaps 26;

QY 20 VSP-VTPIAPAEETGTTITVQDTQKATYKAYKVFAEDINAVNSD-----SNKGASYL 73  
Db 580 IAPIVTSILRNPTSEITNADS---LTFQV--TFSENQVNVNDVDFILSGTGVSGATISN 634

QY 74 PQGKEAEYKASDFRNSLFTTTNG--GRYTVTKDTASANEIATWAKSISANTTP----- 126  
Db 635 VSGS-----GSVTIVTVTGIANNGVTNLDFAEQNIIRDANNAALSNTTPTTDEQ 684

QY 127 -----VSTVBSNNDGTEVINVSQYGYVYSSTVNNGAVIMVTSV-----TP 168  
Db 685 YTLDNTPAASITLTDANITADDIINAE-----SGQAIPIITGVGGEFNVGDIV 733

QY 169 NATIHEKNTDATWGGG-----GKTVDOKTYTSV 196  
Db 734 TLTVDNKTFTGAVGAGGLFSINVPGSDLIVDADLTIAASIAATTDAAGNLGSAATDNQTYTV 793

QY 197 GDTVXYTITYKNVNYHGTKEVYQYVVKDTPSASVDLNEGSYEVTTD-----GSGN 250  
Db 794 DTTAPIITVNDVT-----ADNIIAASGGQAIPIITGVGGEFNVGD 836

QY 251 ITTLTQSGSEKATGKYNLLENNFTIITIPWAAATNPTGNTONGANDDFFYKGINTITVTV 310  
Db 837 TVTLTVNGKPTFG---TVDANGDFSIDV-----LGDDLNGSD-----LTIAA 876

QY 311 T-GVLKSKAGPSADLPENTNIATNPNTSNDPQKV---TVRDGQITIKIDGS 351  
Db 877 SVATTTDAAGNPGSADNQTYYTVDTTAPTNTIAIAVDDIINAIVEAGSPVAVSGTTTGVED 936

QY 352 GQITIKKIDGST-KASLOGAIFVLKNATGQFLNFDNTNVE-----MCTEANATEYTTG 404  
Db 937 GQVTVTIDGNTYATVTVGNATFNIPVADIANFEATEEVATVSDLAGNPATPATRNIT 996

QY 405 ADGI---ITITGLKEGYTYLVEKKAPLGYNLLDNSQKVLGDGATD-TTNSDNLNVPV 460  
Db 997 VDTVAPAVTIDSISDDT-----GAQANDFITNDDTLVFNGTA 1033

QY 461 ENNKGTLPSTGGIGTFTTPIIGAILVTIGAG 491  
Db 1034 EAD-STVVVSLDIE-----IGTVANGAG 1057

RESULT 28  
T28680  
fibrinogen-binding protein homolog - Staphylococcus aureus  
C:Species: Staphylococcus aureus  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C:Accession: T28680  
R:Josefsson, E.; McCrea, K.; Ni Eidhin, D.; O'Connell, D.; Cox, J.; Hook, M.; Foster, T.  
Microbiology 144, 3387-3395, 1998  
A:Title: Three new members of the serine-aspartate repeat protein multigene family of St  
A:Reference number: Z20510; MUID:99098700; PMID:9884231  
A:Accession: T28680  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1166 <JOS>



A;Cross-references: EMBL:AJ005647; NID:el1318793; PID:c1318794; PIDN:CAA06652.1  
C;Genetics:  
A;Gene: sdrE

Query Match 7.0%; Score 180; DB 2; Length 1166;  
Best Local Similarity 23.3%; Pred. No. 0.04;  
Matches 115; Conservative 70; Mismatches 185; Indels 124; Gaps 26;

QY 35 ITVQDTOKGATYKAYKVFDAEIDN-ANVSDSNKDGASYLIPQGEAE-YKASTD-----F 87  
DB 452 IYVNPDKSATNTKVDIAGSDVDYGNKILGN---GSTIIDQNTKIKVYKNSDQQLPOS 508  
QY 88 NSLFTTTTNGGRTVTKKDTASANEIAT-----WAKSISANTTPVSTVSTESNNDGT 138  
DB 509 NRIYDFQYEDVTSQFDPNKKSFNNVATLDFGINSAYIIKVKVSKYPTSDGELDAQGT 568  
QY 139 EVINVSQYGYVSSVYNNGAVIMVTSVTPNATIEHKNTDATWGDGGGKTVDQKTSYSGD 198  
DB 569 SMRTTDKYGYYVAGYSN---FVTS-----NDTGGDGTVPK-EEKLYKIGD 612  
QY 199 TVKTYITYKNVNVHGTKEKYYQYVYKDTMPSASVVDLNEGSYEVTTDGSNITLTQGS 258  
DB 613 YWVEDVDKGVQGTDSKEK-----PMANVL-----VTLFYPDG--TTKSVRT 652  
QY 259 EKATGKYNL--LENNNFITIPWAAATNTPTG---NTQNGANDDFYKGINITITVTVGV 313  
DB 653 D-ANGHVEFGGLKDEGYTVKF-----ETGTGYLPTKVNGTTDGERKDSNGSSVTKINGK 706  
QY 314 LKSAKPGSADLP-----ENTNIATINPNTSNDPQGVTVRQGITIK-----KID 360  
DB 707 DMSLDTGFYKEPKYNLGDYVWEDTKDGIQ---DANEFG---IKDKVTLKDSGTGKI 759  
QY 361 GSTKASLQGAIFVLKATQFLNFNDTNN---VEWGTEA-----NATEYTTGADGII 409  
DB 760 GTTTT-----DASGKY-KPTDLONGYVVEFTPAGYTFPVKNITADDDKDSNGLT 808  
QY 410 TITLKEGTYLYVEK-----KAPLGYNLLDNSQKVLGDGATDTTNSDNLNVPTVENN 463  
DB 809 TTGVIKDADNWLDRGFYKPKYSLGYPVWYDSNK-----DGKQDSIEKGIKDVTVTLQNE 864  
QY 464 KGTPLSTGGIGTT 477  
DB 865 KGEV-----IGTT 872

RESULT 29  
T28681  
rib protein - Streptococcus agalactiae  
C;Species: Streptococcus agalactiae  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C;Accession: T28681  
R;Wastfelt, M.; Stalhammar-Carlismalm, M.; Delisse, A.M.; Cabezon, T.; Lindahl, G.  
J. Biol. Chem. 271, 18932-18937, 1996  
A;Title: Identification of a family of streptococcal surface proteins with extremely rep  
A;Reference number: 220511; MUID:96324974; PMID:8702550  
A;Accession: T28681  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1231 <WAS>  
A;Cross-references: EMBL:U50333; NID:gl620647; PID:gl620648; PIDN:AAC44468.1  
C;Genetics:  
A;Note: rib

Query Match 7.0%; Score 180; DB 2; Length 1231;  
Best Local Similarity 21.9%; Pred. No. 0.043;  
Matches 127; Conservative 70; Mismatches 200; Indels 183; Gaps 31;

QY 13 SLAFGMVSPVPTAFAAETGTTVQDTQKATYKAYKVFDAEID-----NANV 61  
DB 740 TVAFETPVDATPGDKPAKV-VVYTPDGSKDTVDVTVKVVDPRTDADKNDPAGKQVNV 798  
QY 62 SDSNK--DGASYL--IFQGEAEYKASTDFNSLFTTTTNGR-----TYVT-KKDT--- 107

DB 799 GETPKAEDSIGNLPDLPGKTTVAFETPVD-----TATPGDKPAKVVVVTPDGSKDTVDV 852  
QY 108 -----ASANEIATWAKSISANTTPVSTVSTESN-----NDGTEV-----INTVSQYCY 148  
DB 853 TVKVVDPRTDADKNDPAGKQVNVGNETPKAEDSIGNLPDLPGKTTVAFETPVDATPG- 911  
QY 149 YVSSVTVNGGAVIMVT-----SVTPNATIEHKNTDATWGDGGGKTVDQKTSYSGDVTVK 201  
DB 912 -----DKPAKVVVVTPDGSKDTVDVTVKVVDPRTDADKNDPAGK--DQOV-NVGETPK 961  
QY 202 YITYTKAVNY-HGTEKVVQYVVKDTMPS---ASVVDLNEGSYEVTTDGSNITLTQGS 257  
DB 962 AEDSIGNLPDLPGKTTVAFETPVDATPGDKPAKV-----VTYPDGSKDTVDVTDV 1011  
QY 258 SEKATGKYNLLENNNFITIPWAAATNTPTGNTONGANDDFYKGINITITVTVGLKSG 317  
DB 1012 -----VKVVDPRTD-----ADKNDPAGKQDQ-----VNVG 1036  
QY 318 AKPGS-----ADLPENTNIATINPNTSNDPQGVK-----VTVRDG-----QITIKKID 360  
DB 1037 ETPKAEDSIGNLPDLPGKTTVAFETP-VDTATPGDKPAKVVVVTPDGSKDTVDVTVKVV 1095  
QY 361 GSTKASLQGAIFVLKN---ATQQLNFNDTNNVEWGTEANATEYTTGADGIIITGLKSG 417  
DB 1096 PRTDAD-----KNDPAGKQVNVGNETPKAE-----DSIGNLPDLPGK 1133  
QY 418 TYLVLE---KKAPLGYNLLDNSQKVL--GDGATDTTNSDNLNVNPTVE----- 461  
DB 1134 TVAFETPVDATPG-----DKPAKVVVVTPDGSKDTVDVTVKVVDPRTDADKNDPAGKQ 1189  
QY 462 --NNKTELPSTGGIGTTIFYIIGAILVIGAGIVLVARR 499  
DB 1190 QVNGKGNLPGATGNATPFNVVALLTIMSSVGLLSVSKKK 1229

RESULT 30  
S28634  
adhesin AIDA-I precursor - Escherichia coli plasmid pIB6  
C;Species: Escherichia coli  
C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999  
C;Accession: S28634; S22680; S28881; S72657  
R;Benz, I.  
submitted to the EMBL Data Library, March 1992  
A;Reference number: S28634  
A;Accession: S28634  
A;Molecule type: DNA  
A;Residues: 1-1286 <BEN>  
A;Cross-references: EMBL:X65022; NID:g42254; PIDN:CAA46156.1; PID:g42255  
R;Benz, I.; Schmidt, M.A.  
Mol. Microbiol. 6, 1539-1546, 1992  
A;Title: AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic Escheric  
A;Reference number: S22680; MUID:92326638; PMID:1625582  
A;Accession: S22680  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 839-1286 <BE2>  
A;Cross-references: EMBL:X65022  
A;Experimental source: strain 2787  
A;Accession: S28881  
A;Molecule type: protein  
A;Residues: 50-56 <BE3>  
R;Suhr, M.; Benz, I.; Schmidt, M.A.  
Mol. Microbiol. 22, 31-42, 1996  
A;Title: Processing of the AIDA-I precursor: removal of AIDA and evidence for the outer m  
A;Reference number: S72657; MUID:97055419; PMID:8899706  
A;Accession: S72657  
A;Molecule type: protein  
A;Residues: 847-856 <SUH>  
A;Experimental source: DAEC strain 2787  
C;Genetics:  
A;Keywords: plasmid pIB6  
C;Keywords: membrane protein



F:1-49/Domain: signal sequence #status predicted <SIG>  
F:50-1286/Product: adhesin AIDA-I #status predicted <MAT>

```
Query Match      7.0%; Score 180; DB 2; Length 1286;
Best Local Similarity 20.6%; Pred. No. 0.046;
Matches 125; Conservative 76; Mismatches 185; Indels 220; Gaps 30;

Qy 51 VFDAEI-DNANVSDSNKDGASYLIPQGEAEYKASTDFNS-----LFTTTTNG 97
Db 164 IFSGGITDSTNIS-----SGQQRVSGGVA---SNTTINSSGAQNILSEGAISTHSSG 216

Qy 98 GRTVVTKKDTASANEIATWAKS-----ISANTTPVSTVTSNNDGTETVINVSQYGYVYS 152
Db 217 GNQYI-----SAGANATEIIVNSGFORVNSGAVATGTVL---SGGTQ--NVSSGSAIST 267

Qy 153 STVNGAVIMVTSVTPNATIHKNITDATMGDGGGKTVD-----QKTYSVG 197
Db 268 SVYNSG---VQTVFAGATV---TDTTVNSGGNQNISSGGIVSETTVNVSGTQNIYSGG 319

Qy 198 DTVKYITYKNAVNYHGT-----EKVQYVVKDTPSASV-----VDLNEGSY-EVT 243
Db 320 SALSANIKGSQIVNSEGTAINTLVSDGGYQHIRNGGIASCTIVNOSGYVNISSGYAEST 379

Qy 244 ITDGSNITTLTQSGEKAT-----GKYNLLENNFTITIPWAATNTPTGNTQNGANDDF 298
Db 380 IINSGGTLRVLDGYARGTILNNSGRENVSNGVSYNAMI-----NT--GGNQYI 427

Qy 299 FYKG-----INTI-----TVTYTGVLSKAKPGSADLPEN----- 328
Db 428 YSDGEATAAIVNTSGFORINSGGTAPVQNSVVVTRIVSSAAKPFDAEVYSGGQTVYLWR 487

Qy 329 -----TNIATINPT-----SNDDPGQKVTVRDQI--TIKKIDG-----STK 364
Db 488 GIWYSNFLTAVMSMFPCTASGANVNLGRLNAPAGNVVGTILNQEGRQYVYSGATATSTV 547

Qy 365 ASLQGAIFVLKN--ATCQFLNFNDTNVVEGTANATEYTTGA-----DGIITITGLXE 416
Db 548 GNNEGREYVLSGGITDGTVLNSGGLQAVSSGGKASATVINEGGAQFVYDGGQVGTGNIKN 607

Qy 417 GTYVLVEKKAPL-----GYNLL-----DNSQKVLGDG----- 444
Db 608 GGTIRVDSGASALNIALSSGGLFTSTGATLPCLTTMAALSVSQNHASNIVLENGGLLRV 667

Qy 445 -----ATDIT-----NSDNLNVNPTVENKNG-----T 466
Db 668 TSGGTATDTVNSAGRLRIDGGCTINGTTTINADGIVAGTNIQNDGNFILNLAENYDFET 727

Qy 467 ELPSTG 472
Db 728 ELSGSG 733
```

Search completed: July 22, 2004, 10:29:19  
Job time : 24 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 09:59:46 ; Search time 13 Seconds  
(without alignments)  
2010.709 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKKWISQLLVASLAFGMAV.....GAILVIGAGIVLVARRRLLRS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	284	11.0	537	1	TEE6_STRPY
2	258	10.0	533	1	FM1_ACTVI
3	231.5	9.0	534	1	FM2_ACTNA
4	185.5	7.2	1025	1	SLAP_CAUCR
5	180	7.0	1286	1	AIDA_ECOLI
6	179.5	7.0	2021	1	OMPA_RICCN
7	178	6.9	1041	1	EGT2_YEAST
8	177.5	6.9	1183	1	CNA_STAU
9	176	6.8	1276	1	PMP2_CHLPN
10	169	6.6	2249	1	OMPA_RICRI
11	167.5	6.5	939	1	EAE_ECO27
12	167.5	6.5	1902	1	P1P_LACIC
13	167	6.5	2003	1	YDBA_ECOLI
14	165.5	6.4	1902	1	P2P_LACIC
15	165	6.4	1643	1	OMPE_RICPR
16	165	6.4	1655	1	OMPB_RICCN
17	164.5	6.4	1902	1	P3P_LACIC
18	164	6.4	1569	1	YPUA_ECOLI
19	164	6.4	1953	1	BIGA_SALTY
20	162.5	6.3	1902	1	P2P_LACPA
21	159	6.2	2334	1	WAPA_BACSU
22	158	6.1	1004	1	SLPO_BACBR
23	158	6.1	1039	1	AG43_ECOLI
24	158	6.1	1228	1	SLAP_BACST
25	157.5	6.1	670	1	VG50_HSV11
26	157.5	6.1	1656	1	OMPB_RICJA
27	157	6.1	439	1	SLAP_LACHE
28	156.5	6.1	666	1	MUR2_ENTHR
29	156.5	6.1	1020	1	BCA_STRAG
30	156.5	6.1	1861	1	APU_THETU
31	156.5	6.1	2358	1	YEEJ_ECOLI
32	155	6.0	2680	1	YEEJ_ECO57
33	153.5	6.0	497	1	FLIC_ECOLI

#### ALIGNMENTS

#### RESULT 1

TEE6\_STRPY STANDARD; PRT; 537 AA.

AC P18481;

DT 01-NOV-1990 (Rel. 16, Last Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Trypsin-resistant surface T6 protein precursor (T6 antigen).

GN TEE6

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

34	153.5	6.0	1300	1	120K_RICRI
35	153.5	6.0	1554	1	OMPB_RICRI
36	152.5	5.9	934	1	EAE_ECO57
37	152.5	5.9	939	1	SLAP_CAMPE
38	149	5.8	936	1	PMP7_CHLPN
39	148.5	5.8	490	1	TA53_TREDE
40	148.5	5.8	1075	1	FLO5_YEAST
41	148	5.7	738	1	Y013_BPL2
42	148	5.7	1250	1	YFAL_ECOLI
43	148	5.7	1256	1	MRP_STRSU
44	147.5	5.7	465	1	SLAP_LACBR
45	147	5.7	1460	1	PMP_CHLMU
46	145	5.6	1325	1	YDBK_ECOLI
47	144.5	5.6	863	1	YEO_ECOLI
48	144.5	5.6	1176	1	SLAP_BACSH
49	144	5.6	670	1	YFGG_SCHPO
50	144	5.6	1848	1	CBPA_CLOCL
51	143.5	5.6	1645	1	OMPB_RICTY
52	143.5	5.6	1672	1	PMPB_CHLMU
53	143.5	5.6	5703	1	MUSB_HUMAN
54	143	5.5	918	1	YMOB_CAEEL
55	143	5.5	935	1	EAE_ECO11
56	143	5.5	1967	1	CD87_DROME
57	143	5.5	3178	1	YS89_CAEEL
58	142.5	5.5	806	1	SURV_BACSU
59	142.5	5.5	1754	1	PMPB_CHLTR
60	141.5	5.5	762	1	SLAP_ACEKI
61	141.5	5.5	1065	1	SED4_YEAST
62	141.5	5.5	1289	1	VG34_BPT4
63	141	5.4	2710	1	TOXA_CLODI
64	140.5	5.4	881	1	PRY3_YEAST
65	140.5	5.4	1186	1	CEAA_BACTS
66	140	5.4	313	1	KRE1_YEAST
67	140	5.4	444	1	SLAP_LACAC
68	140	5.4	827	1	CSG_HALVO
69	140	5.4	1628	1	NAGH_CLOPE
70	139.5	5.4	1243	1	VG37_BPK3
71	139.5	5.4	1419	1	ALAI_CANAL
72	139	5.4	1537	1	FLO1_YEAST
73	139	5.4	2432	1	Y43R_IRV6
74	138.5	5.4	1256	1	ATL_STAAU
75	138	5.4	880	1	LYTD_BACSU
76	136.5	5.3	527	1	VG12_BPT4
77	136.5	5.3	1036	1	HP12_DEIRA
78	136.5	5.3	1258	1	ICEN_ERWHE
79	136	5.3	905	1	HXA1_HAEIN
80	136	5.3	1853	1	CIPA_CLOTM
81	136	5.3	2602	1	FLNB_HUMAN
82	135.5	5.3	466	1	FLID_SALTY
83	135	5.2	995	1	YI09_YEAST
84	134.5	5.2	754	1	PLE3_CANAL
85	134	5.2	1778	1	N189_SCHPO
86	133.5	5.2	930	1	PMP8_CHLPN
87	133.5	5.2	1018	1	FNBA_STAAU
88	133.5	5.2	1034	1	ICEN_PANAN
89	133.5	5.2	1150	1	APMU_PIG
90	133.5	5.2	1562	1	SPAP_STRMU

P14914	rickettsia
Q53047	r outer mem
P43261	escherichia
P35827	campylobact
Q92898	chlamydia p
P18164	treponema d
P38894	saccharomyc
P42548	bacterioph
P45508	escherichia
P32653	streptococc
Q05044	lactobacill
Q9PJY1	chlamydia m
P32051	escherichia
P33924	escherichia
P38537	bacillus sp
O13854	schizosacch
P38058	clostridium
P96989	r outer mem
Q9PJY2	chlamydia m
Q9H84	homo sapien
P34487	caenorhabdi
O31000	escherichia
Q9VGS5	drosophila
Q96824	caenorhabdi
P29141	bacillus eu
O84418	chlamydia t
P22558	acetogenium
P25365	saccharomyc
P18771	bacterioph
P16154	clostridium
P26831	clostridium
P47033	saccharomyc
Q45710	bacillus th
P17260	saccharomyc
P35829	lactobacill
P25062	halobacteri
Q38394	bacterioph
O13368	candida alb
P32768	saccharomyc
P18305	chilo iride
P52081	staphylococ
P39848	bacillus eu
P10930	bacterioph
P13126	deinococcus
P16239	erwinia her
P44602	haemophilus
Q06851	clostridium
Q75369	homo sapien
P16328	salmonella
P40442	saccharomyc
Q9UVX1	candida alb
Q9UTK4	schizosacch
Q92393	chlamydia p
P14738	staphylococ
P40879	pantoea ana
P12021	sus scrofa
P23504	streptococc

OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=D471 / Serotype M6;  
 RX MEDLINE=90264329; PubMed=2189957;  
 RA Schneewind O., Jones K.F., Fischetti V.A.;  
 RT "Sequence and structural characteristics of the trypsin-resistant T6  
 surface protein of group A streptococci";  
 RL J. Bacteriol. 172:3310-3317(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21843101; PubMed=11854196;  
 RA Bessen D.E., Kalia A.;  
 RT "Genomic localization of a T serotype locus to a recombinatorial zone  
 encoding extracellular matrix-binding proteins in Streptococcus  
 pyogenes";  
 RL Infect. Immun. 70:1159-1167(2002).  
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 an amide bond (Potential).  
 CC -1- MISCELLANEOUS: MORE THAN 25 SEROLOGICALLY DIFFERENT T ANTIGENS ARE  
 FOUND TO BE EXPRESSED INDEPENDENTLY OF M PROTEIN, BUT CERTAIN T  
 PROTEINS OCCUR ONLY IN ASSOCIATION WITH SPECIFIC M SEROTYPES  
 (I.E., T6 PROTEIN IS ALWAYS FOUND ON M6 STREPTOCOCCAL STRAINS).  
 CC -1- MISCELLANEOUS: THE T PROTEIN IS EXPOSED ON THE CELL SURFACE WHEN  
 FIBRILLAR M PROTEIN IS ENZYMATICALLY REMOVED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M32978; AAA27019.1; -;  
 DR EMBL; AY049087; AAL11467.1; -;  
 DR PIR; A35400; A35400.  
 DR InterPro; IPR008454; Cna\_B.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR Pfam; PF05738; Cna\_B; 1.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PROSITE; PS0847; GRAM\_POS ANCHORING; 1.  
 KW Cell wall; Peptidoglycan-anchor; Antigen; signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 507 TRYPsin-RESISTANT SURFACE T6 PROTEIN.  
 FT PROPEP 508 537 REMOVED BY SORTASE (POTENTIAL).  
 FT DOMAIN 310 330 HYDROPHILIC.  
 FT SITE 504 508 LPXTG SORTING SIGNAL (POTENTIAL).  
 FT MOD RES 507 507 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
 SQ SEQUENCE 537 AA; 57674 MW; 8A008A6240998A5 CRC64;  
 Query Match 11.08; Score 284; DB 1; Length 537;  
 Best Local Similarity 22.28; Pred. No. 8.6e-09;  
 Matches 135; Conservative 89; Mismatches 195; Indels 188; Gaps 28;  
 QY 10 LVASLAF-----GMAVSPVPTAFAAETGTTIVQDTQKATYKAYKVPDABIDNANVSDSN 65  
 DB 1 MLACLAILAVVLGMLTRVSALS-KDDTAQLKLTNIEGGPTVLYKIGEGVY-----N 51  
 QY 66 KDGASYLIPOKAEAYKASTD---FNSLFTTTNGRTGYTK-----KDTASANEIATW-A 117  
 DB 52 TNGDSFI--NPKYAEVSLTETGTSQETITTIANGINTGKIKPFSTENVISINGTATYNA 109  
 QY 118 KSIISA-----NTTPVSTVTSNNDDTEVI-NV-SOYGYVYVSVTVNNGAVIMV 163  
 DB 110 RGA SVYALTATGDTGTYNPILLAA SYNGEGLNLTNIDSKSNLYLQGTISVAKSSLP 169  
 QY 164 TSVPNPNATIEKNTDATWGDGGKTVQDKYVSVDGTVKTYTY-----KNVNNYHGT 217  
 DB 170 T-----KKVTGTDID-----VNKKTSLGSLVLSYSLTPELPSTYKAVN----KT 210  
 QY 218 VYQVVIKDTWPSASVDLNEGSYEVTTIDSGNTITTTQSEKATGKYNLLENNFTIT 277

Db 211 VY---VSDNMSEGLTENFN-----SLIV-EWKGGMANITDG-----SVVVENTKIGI- 254  
 QY 278 IPWAAINTPTGNTONGANDDFYKGINITIT--VTYTGVLKSGAKPGSADLPENTN-IATI 334  
 Db 255 -----AKEVNGFNLSFYDLSIESIPNISYKAVVNNKAIVGEENPKABFFYSN 305  
 QY 335 NPWTSN-----DBPGQKVTVRDG-----QITIKKIDGSTKASLOCAIVLKNATG 379  
 Db 306 NPTKGYTNDLDRKPKNGITSKEDSKIVYVYQIAFRKVDVSVKTPFLGAIFGVVDTSN 365  
 QY 380 QFLNFNDTNNVEMWTEANATEYTTGADGIITITGLKEGYLYLVEKCAPLGYNL----- 432  
 Db 366 KLIDI-----VTTNKGVAISTQVSSGKYKIKELKAPKGYSLNTETYEI 409  
 QY 433 -----LONSQKV----- 439  
 Db 410 TANWVATVKTSAKSTTYSKDKATDSEQVGLMKNGIFYSIDSRPTGNDVKDAYIE 469  
 QY 440 ---ILGDGAT---DTTNSDNLVNPVTENNKGTELPSTGGTIGTTFIVFIIGAILVIGA-GI 492  
 Db 470 STKALTDGTTFSKNSGSGTVLLETIDIPNKLGLPELSTGSGITLYLFAIGSAAMIGAIGI 529  
 QY 493 VLVARRR 499  
 Db 530 YIVKRRK 536  
 RESULT 2  
 FM1 ACTVI  
 ID NCBI\_TaxID=1314; PRT; 533 AA.  
 AC P18477;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Fimbrial subunit type 1 precursor.  
 OS Actinomyces viscosus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Actinomycineae; Actinomycetaceae; Actinomycetes.  
 OX NCBI\_TaxID=1656;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-56.  
 RX STRAIN=TL4V;  
 RC MEDLINE=90236904; PubMed=1970561;  
 RA Yeung M.K., Cisar J.O.;  
 RT "Sequence homology between the subunits of two immunologically and  
 functionally distinct types of fimbriae of Actinomyces spp.";  
 RL J. Bacteriol. 172:2462-2468(1990).  
 CC -1- FUNCTION: MAJOR FIMBRIAL SUBUNIT OF ACTINOMYCES VISCOSUS.  
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 an amide bond (Potential).  
 CC -1- SIMILARITY: TO FIMBRIAL SUBUNIT 2 OF ACTINOMYCES NAESLUNDII.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M32067; AAA62572.1; -;  
 DR PIR; A35259; A35259.  
 DR InterPro; IPR008454; Cna\_B.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR Pfam; PF05738; Cna\_B; 2.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PROSITE; PS0847; GRAM\_POS ANCHORING; 1.  
 KW Fimbria; Cell wall; Peptidoglycan-anchor; signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 499 FIMBRIAL SUBUNIT TYPE 1.  
 FT PROPEP 500 533 REMOVED BY SORTASE (POTENTIAL).  
 FT SITE 496 500 LPXTG SORTING SIGNAL (POTENTIAL).  
 FT SITE 496

```
FT MOD_RES 499 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 533 AA; 56899 MW; F00299559C702FC4 CRC64;

Query Match
Best Local Similarity 10.0%; Score 258; DB 1; Length 533;
Matches 146; Conservative 67; Mismatches 223; Indels 126; Gaps 30;

QY 13 SLAFGMAVSPVPIAFAETGTHITVDQTKGATYKAYKVFDAIDNANVSDNSKDGASVL 72
DB 17 TLAAAGALVAP-TGAAPADPNGSTI-DPDAATLTTHKCEQDTNGVREGTGNED----- 69
QY 73 IQQKREAEKASTDFNSLFT-----TTNGRTTV-TKDTASANEI-ATWAKISANT 124
DB 70 -PQ---AECKPVSDFEFTIKLNVLTYYDGHKTLADLKGDDVVKAGALKSTTVQKITCA 125
QY 125 TPVSTVSTNDSNGTEVINVSQYGYVSVSTVNGAVI-----MVT--SVTP-----N 169
DB 126 NGLASFDTAQ---TEV-----GAVLVSETRPDVKVPAEDFVVLPMNPQDTAKWYN 176
QY 170 ATIHENKNTDATWGDGGKTV-DQKYSVGDVTKYITITYK-NAVNHGTEKVVQYVVKDTM 227
DB 177 VHVYPKNTLS-----GVDKQVTDKPAKSGRDTITYTITSIPKVDYPGGARIKRYEVDRL 232
QY 228 -----PSASVVDLMEGSEVETITDGSNITTLTQSEKATGKYNLLNENNFIT 277
DB 233 DKRIKKEALTVPVKIV-----GQNEVTLAB-TTDTYLTIT-----AEKG-----DHNAWIIQ 277
QY 278 IPWAATNPTGNTGNDANDFFYKGINT---ITVTYTVGLKSGAKPGSADLPENTNIAFI 334
DB 278 LTEEGRRKASEARYNGETKQLVTLNAKFAAVALNGLDLSNTAGLIPNDSP-NFTWDPN 336
QY 335 NPNTSNDPDG---QKTVRDGQITTKI---DGSKASLOQAI FVL-----KNATQGLNF 384
DB 337 NPGTTTIDIPGTPPVLKSYGKVLTKTGTDDLADKTKYNGAQFOVECTKTASGATLRD 396
QY 385 ND--TNVWEGTEANATEVTTGADGIITITGLK-----EGTYLVLEKAP 427
DB 397 SPDSQTVDPLTIGGEKTEFTAGQGVVEINLRANDYVNGAKKQDLTDEYVCLVETKAP 456
QY 428 LGYN-----LLDNSQKVLGDGATDTTNSNLLNPTVNNKGTPELSTGGIGTT 477
DB 457 EGYNLQADPLPRVLAKEAK-----KAATEVTVTD-----IPKNAGFRLPLTGANGVI 505
QY 478 IFYIIGAILVIGAGIVLVARR 499
DB 506 FLTIAGALLVAGGAVVAYANKR 527

RESULT 3
FM2_ACTNA STANDARD; PRT; 534 AA.
AC P12616;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fimbral subunit type 2 precursor.
OS Actinomyces naeslundii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Actinomycineae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=VVU45;
RA MEDLINE=88314866; PubMed=2900829;
RA Yeung M.K., Cisar J.O.;
RT "Cloning and nucleotide sequence of a gene for Actinomyces naeslundii
RT VVU45 type 2 fimbriae.";
RL J. Bacteriol. 170:3803-3809(1988).
CC -!- FUNCTION: MAJOR FIMBRIAL SUBUNIT OF ACTINOMYCES NAESLUNDII.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -!- SIMILARITY: TO FIMBRIAL SUBUNIT 1 OF ACTINOMYCES VISCOSUS.
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M21976; AAA21931.1; -.
DR PIR; A32347; A32347.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B_2.
DR TIGFAMS; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
KW Fimbria; Cell wall; Peptidoglycan-anchor; Signal.
FT SIGNAL 1 32
FT CHAIN 33 495 FIMBRIAL SUBUNIT TYPE 2.
FT PROPEP 496 534 REMOVED BY SORTASE (POTENTIAL).
FT SITE 492 496 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 495 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 534 AA; 56574 MW; 6C56C3AB493D5751 CRC64;

Query Match
Best Local Similarity 9.0%; Score 231.5; DB 1; Length 534;
Matches 141; Conservative 52; Mismatches 202; Indels 155; Gaps 28;

QY 54 AEIDNANVSDSNK-----DGASVLIPOGKEA--EYKASTDFNSLFTTTTNG 97
DB 30 AQAEANAHGDIINTEALGSLTIHKHLNGDNIGAPDGTASNDGKAPVSGVQFYATYEIN 89
QY 98 GRVYVTKDTSANEIATWA---KSISAN---TTPVSTVSTES-----NNDGTEVINVS 144
DB 90 GIDLKTSQWAKVNAITNGAIPDNACANCPGQTLPNVTFRRSSRVSGDTRDGEAKIESL 149
QY 145 QGYHYVSVSTVNGAVI-----MVTSTVTPNATHEKNTDATW-----GDGSK 187
DB 150 PVKALVLCETTPGNIQVQAKPFVVTIPIHPNTA---AKADGTWLYDVHVPKNEKIEVAK 206
QY 188 TV-DQKT--YSVGDTPVKY---TITYKNAVNYHGTQVYVVKDTMPSASVVDLNEGSY 240
DB 207 TIEDORNGYIVGSKVRFVSVSTPLKLDNSY---KYQF---KUT-----LDNRLLK 253
QY 241 EVTITDGSNITTLTQSEKATGKYNLLNENNFITIPWAATNPTGNTGNDANDFFY 300
DB 254 QVTATDVTLGGLTDLDEGTD---YTLGTDGQTVTVTF-----NQGLSK---L 294
QY 301 KGI--NTITVTVTVGLK---SGAKPGSADLPENTNIAINP-----NTSNDPDGQKVT 348
DB 295 KGNPGQKLQAVPEGVVSEVGDGSIINNTAQLISDTTYAEQPPAPETPPANPDNPPTTEQVT 354
QY 349 VRDGOITIKKIDGS---TKASLOQAI FVL-----KNATGQFLNFDNTNNVSW 392
DB 355 SKWGLTLTIKKVDGNDNRSGDKGLKGAEFQIYKAKDAYADTCSPEADGQPLTN----- 407
QY 393 GTEANATEYTTGADGIITITGL-----KEGYVLVLEKAPGLNLLDN 435
DB 408 ---GESFTTTEGGGTINFKALFVSDSVQDTRGNRVDPAPHRCYVLVETKAPAGVVLPA 463
QY 436 SQKVL---GDGATDTTNSDNLNLYNPTVNNKGTPELSTGGIGTTFIYIIG-AILVIGAG 491
DB 464 ASRAITVEFGAVTQQVVVIDN--VKQSVPG-----LPLTGANGMLILTASGAALLMIAVG 516
QY 492 IVLVARRRLR 501
DB 517 SVLVARYRER 526

RESULT 4
SLAP_CAUCR
ID SLAP_CAUCR STANDARD; PRT; 1025 AA.
AC P35828; Q46015; Q9RF12;
```

01-JUN-1994 (Rel. 29, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 10-OCT-2003 (Rel. 42, Last annotation update)  
 S-layer protein (Paracrystalline surface layer protein).  
 R8AA OR CC1007.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
 OC Caulobacteraceae; Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=93007489; PubMed=1393820;  
 RA Gilchrist A., Fisher J.A., Smit J.K.;  
 RT "Nucleotide sequence analysis of the gene encoding the Caulobacter  
 crescentus paracrystalline surface layer protein.";  
 RL Can. J. Microbiol. 38:193-202(1992).  
 RN [2]  
 RP REVISIONS TO 376; 636 AND 842-843.  
 RA Awram P.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JS3001;  
 RA Bingle W.H., Awram P.A., Nomellini J.F., Smit J.K.;  
 RL "The secretion signal of C. crescentus S-layer protein is located in  
 the C-terminal 82 amino acids of the molecule.";  
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Petočka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.G., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Uterback T., Tran K., Wolf A., Vamathevan J., Emolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RL "Complete genome sequence of Caulobacter crescentus.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 RN [5]  
 RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=89008089; PubMed=3049545;  
 RA Fisher J.A., Smit J.K., Agabian N.;  
 RT "Transcriptional analysis of the major surface array gene of  
 Caulobacter crescentus.";  
 RL J. Bacteriol. 170:4706-4713(1988).  
 RN [6]  
 RP CHARACTERIZATION.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=98292737; PubMed=9620954;  
 RA Awram P., Smit J.K.;  
 RT "The Caulobacter crescentus paracrystalline S-layer protein is  
 secreted by an ABC transporter (type I) secretion apparatus.";  
 RL J. Bacteriol. 180:3062-3069(1998).  
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
 OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A  
 PHYSICAL BARRIER TO PARASITES AND LYtic ENZYMES.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER  
 (TYPE I) SECRETION APPARATUS.  
 CC -1- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE  
 SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A  
 SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC EMBL; AF062345; AAC38665.2; --  
 CC EMBL; AF193063; AAF19365.1; --  
 DR EMBL; AB005779; AAK22991.1; ALT\_INIT.  
 DR PIR; A48995; A48995.  
 DR HSSP; P22629; LSWC.  
 DR TIGR; CC1007; --  
 DR InterPro; IPR001343; Hemlysn\_Ca\_bind.  
 DR Pfam; PF00353; hemolysinCBind; 3.  
 DR PRINTS; PRO0313; CABNDNGRPT.  
 KW Cell wall; S-layer; Calcium-binding; Complete proteome.  
 FT INIT MET 0  
 SQ SEQUENCE 1025 AA; 98001 MW; AD7A326E1363D8AC CRC64;  
 Query Match 7.2%; Score 185.5; DB 1; Length 1025;  
 Best Local Similarity 22.1%; Pred. No. 0.0051;  
 Matches 131; Conservative 66; Mismatches 240; Indels 157; Gaps 26;  
 QY 6 IQSLIVASLAFGMVSPVTPPIAFAAETGTTVDQTKGATYKAYKVFDAEIDNANVSDSN 65  
 DB VQAAVTLPTGVTISGIE--TWNVTSGAAITLNTSSGVT-----GLTALNTN 343  
 QY 66 KDGASYLIPQKREAEYKASTDFNSLFTTTNGG-----RTYVTK-KDTASANEIATWAK 118  
 DB TSGAAQTVTAGAGONLTATTAQAANNVAVDGGANVTVASTGVTSGTTTGVANSAAAGTV 403  
 QY 119 SIS-ANTTPVST-----VTESNNDGTEVINVSQYGY 148  
 DB SVSVANSSTTTTGAIAVGTGTAVTVAQTAGNAVNTTLTQADVTVTGNSTTAVTTQTAA 463  
 QY 149 YVVSSTVN---NGAVIMVTSVTPNATIHBEKNTDATWGGGKTVDQKTVSGDVTVKYIT 205  
 DB ATAGATVAGRVNGAVITIDSAASATACKIATVTLGSGAATIDSSALT----- 513  
 QY 206 YKNAVNYHTEK---VYQYVIKDTWPSASVDNLNBSYEVT---ITDGGG---NITTLFQ 256  
 DB ---TVNLSTGTSGLIGRGALTAT-PTANTLTNLVNGLTITGCAITDSEAAADDGPTTNI 569  
 QY 257 GSEKATGKYNLENNNTTITPWAATVPTGNTQNGANDDFYKGINITVT----- 309  
 DB AGTASSTIASLAADATTLNISGDARVITISHTA-----AALTGITVNSVGATL 620  
 QY 310 ----YTGVLKSGAKPGSADLPENTNIA-----TINPNTSNDPQGVTVTRDG-QIT 356  
 DB GAELATGLVFTGGAGADSILLGATTKAIVMGAGDDTTVSSATLGAGSGVGGDGTDLV 680  
 QY 357 KIDGSTKASLOGAIFVLKNATGQF-----LNFNDTNVNVGTEANATEYTT 403  
 DB ANVNGSS-----FSADPAFGFPETLRVAGAAAGSHNANGFTALQLGATAGATTFTN 732  
 QY 404 GADGI-ITITGLKEGTYLVEKKAPLG---YNL-LDNSQKVLGD-----GAT 446  
 DB VAVNGLTVLAAPTGTITVTLANA-TGTSDFVNLTLSSAALAAGTVALAGVETVNIAT 791  
 QY 447 DTTNS---DNLVNPET-----VENNKGTPELSTGGTGTTFYIIGAILVIGAG 491  
 DB DINTTAHVDTLTQATSAKSIVVTGNAGLNTNGTAVTSF---DASAVTGTG 842  
 RESULT 5  
 AIDA\_ECOLI  
 ID AIDA\_ECOLI STANDARD; PRT; 1286 AA.  
 AC Q03155;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Adhesin aida-I precursor.  
 GN AIDA-I.  
 OS Escherichia coli.  
 OG Plasmid pIB6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.  
RC STRAIN=O126:H27 / 2787;  
RX MEDLINE=92326638; PubMed=1625582;  
RA Benz I., Schmidt M.A.;  
RT "AIDA-I, the adhesin involved in diffuse adherence of the  
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is  
RT synthesized via a precursor molecule.";  
RL Mol. Microbiol. 6:1539-1546(1992).  
CC -!- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE  
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI  
CC TO EPITHELIAL CELLS.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL; X65022; CA446156.1; --  
CC PIR; S28634; S28634.  
DR InterPro; IPR006315; Autotransport.  
DR InterPro; IPR005546; Autotransporter.  
DR InterPro; IPR004899; Pertactin.  
DR Pfam; PF03797; Autotransporter; 1.  
DR Pfam; PF03212; Pertactin; 1.  
DR TIGRfam; TIGR01414; autotrans\_bar1; 2. Plasmid.  
KW Cell adhesion; Signal; Outer membrane; Adhesin AIDA-I.  
FT CHAIN 1 49  
FT PROPEP 50 ? ADHESIN AIDA-I.  
FT PROPEP 50 ?  
SQ SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;  
  
Query Match 7.0%; Score 180; DB 1; Length 1286;  
Best Local Similarity 20.6%; Pred. No. 0.013;  
Matches 125; Conservative 76; Mismatches 185; Indels 220; Gaps 30;  
  
QY 51 VFDAEI-DNANVSDNKDGASLYLIPQKEAEYKASTDENS-----LFTTTTNG 97  
Db 164 IISGGITDSTNIS-----SGQQRVSQGVV-----SMTTNSGQAILSEGAISTHSSG 216  
QY 98 GRYYTVTKDASANEIATWAKS-----ISANTPTVSTTESNNDGTEVINVSQYGYVVS 152  
Db 217 GNQYI-----SAGANATETIVNSGQQRVNSGAVATGTVL---SGGTQ--NVSSGSGAIST 267  
QY 153 STVNGAVIMVTSVTPNATIEKNTDATWGDGGKTVD-----QKTYSVG 197  
Db 268 SVYNSG---VQTVFAGATV---TDTTVNSGQNVISGGIVSETTVNSGTONIYSGG 319  
QY 198 DTVKTYITKYNVNYHGT-----EKYQYVIKDTMPSASV-----VDLNEGSY-EVT 243  
Db 320 SALSANIKSQVINSSEGTAINTLVSDGGYQHRNGIASGTIVNSQYGVNISSSGGAEST 379  
QY 244 IYDGSNITTLTQSGSEKAT-----GKYNLLENNNFTTIPWAAATPTGNTQNGANDDF 298  
Db 380 IINSGGTLRLVSDGYARGTILNNSGRENVSNGVSNAMI-----NT--GGNQYI 427  
QY 299 FYKG-----INTI-----TVYTVGLKSGAKPGSADI-PEN----- 328  
Db 428 YSDGEATAIIVNTSGFQRINSGGTAPVQNSVTVVTRTVSSAAKPFDAEYVSGKQTVLWR 487  
QY 329 -----TNIATINPT--SNDPQGVTVRDQI--TIKKIDG-----STK 364  
Db 489 GIWYNFLVAVSMFPGTASGAVNLGSLNFAFAGVNVGTILNQBGQYVYSGATATSTV 547  
QY 365 ASLQGAIFVLKNI--ATGQFLNFNDNNVWEGTEANATEYTCG-----DGIITITGLKE 416  
Db 548 GNNEGREYVLSGIGTDTVLSNGGLQAVSSGSKASATVINEGQAPVVDGGQVGTGNTIKN 607

QY 417 GYYLVVEKKAPL-----GYNLL-----DNSQKVIIGDG----- 444  
Db 608 GGTIRVDSGASALNIALSSGGLFTSTGATLPETLTAALSVSONHASNIVLENGLLRV 667  
QY 445 -----ATDTT-----NSDNLVNPVTVENKKG-----T 466  
Db 668 TSGGTATDTTIVNSAGRLRIDGGTGTGTTTINADGIVAGTINQDNGNFILNLAENYDPET 727  
QY 467 ELPSTG 472  
Db 728 ELSSGS 733  
  
RESULT 6  
OMPA\_RICCN STANDARD; PRT; 2021 AA.  
ID OMPA\_RICCN STANDARD; PRT; 2021 AA.  
AC Q52657; P95591; P95592; P95594; Q52667; Q52668; Q52669;  
AC Q52670; Q52674;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface  
DE antigen) (rOmpA) (rOmp A).  
GN OMPA OR RCL273.  
OS Rickettsia conorii.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=781;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Malish 7;  
RX MEDLINE=94171067; PubMed=8125327;  
RA Croquet-Valdes P.A., Weiss K., Walker D.H.;  
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia  
RT conorii (Malish 7 strain).";  
RL Gene 140:115-119(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Malish 7;  
RX MEDLINE=21442074; PubMed=11557893;  
RA Ogata H., Audic S., Reneato-Audiffren P., Fournier P.-E., Barbe V.,  
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
RA Raoult D.;  
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
RL Science 293:2093-2098(2001).  
RN [3]  
RP SEQUENCE OF 8-204 FROM N.A.  
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;  
RX MEDLINE=97015921; PubMed=8862558;  
RA Roux V., Fournier P.E., Raoult D.;  
RT "Differentiation of spotted fever group rickettsiae by sequencing and  
RT analysis of restriction fragment length polymorphism of PCR-amplified  
RT DNA of the gene encoding the protein rOmpA.";  
RL J. Clin. Microbiol. 34:2058-2065(1996).  
RN [4]  
RP SEQUENCE OF 953-2012 FROM N.A.  
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;  
RA Raoult D., Fournier P.E., Roux V.;  
RT "Phylogenetic analysis of spotted fever group rickettsiae by study  
RT of the outer surface protein rOmpA.";  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-  
CC layer with hexagonal symmetry.  
CC -!- PTM: Glycosylated (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)



```
or send an email to license@sib-sib.ch).
-----
CC ENBL; U01028; AAA17405.1; -
DR ENBL; AE008674; AAL03811.1; -
DR ENBL; U43794; AAB49549.1; -
DR ENBL; U43798; AAB49550.1; -
DR ENBL; U43806; AAB49551.1; -
DR ENBL; U45244; AAB49566.1; -
DR ENBL; U46918; AAB86663.1; -
DR ENBL; U83440; AAC35176.1; -
DR ENBL; U83443; AAC35179.1; -
DR ENBL; U83448; AAC35184.1; -
DR ENBL; U83453; AAC35189.1; -
DR InterPro; IPR06315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRfams; TIGR01414; autotrans_bar1; 1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;
KW Complete proteome.
FT SIGNAL 1 38
FT CHAIN 39 2021
FT DOMAIN 238 946
FT DOMAIN 1424 1528
FT VARIANT 60 60
FT VARIANT 76 76
FT VARIANT 86 137
FT VARIANT 126 133
FT VARIANT 953 954
FT VARIANT 1245 1245
FT VARIANT 1308 1308
FT VARIANT 1877 1877
FT CONFLICT 10 10
FT CONFLICT 92 92
FT CONFLICT 126 126
FT CONFLICT 137 137
FT CONFLICT 157 157
FT CONFLICT 368 369
FT CONFLICT 374 388
FT VARIANT 640 640
FT CONFLICT 669 669
FT CONFLICT 793 793
FT CONFLICT 803 804
FT CONFLICT 809 823
FT CONFLICT 898 898
FT CONFLICT 908 908
FT CONFLICT 985 985
FT CONFLICT 1009 1009
FT CONFLICT 1013 1013
FT CONFLICT 1182 1182
FT CONFLICT 1314 1314
FT CONFLICT 1451 1451
FT CONFLICT 1624 1624
FT CONFLICT 1628 1628
FT CONFLICT 1872 1872
FT CONFLICT 1875 1875
FT CONFLICT 1878 1879
FT CONFLICT 1936 1936
FT CONFLICT 1965 1970
FT CONFLICT 1997 1997
SQ SEQUENCE 2021 AA; 203328 MW; 327FC42D7CB24668 CRC64;
Query Match 7.0%; Score 179.5; DB 1; Length 2021;
Best Local Similarity 23.7%; Pred. No. 0.025;
Matches 143; Conservative 70; Mismatches 227; Indels 163; Gaps 32;
9 LLVASLAFGMVSPV---TPIAFAAETGT-----ITVQDTOKGA-----TYKAY 49
30 MLSSSGALGIAGVGIATNNNAFSDNVGNWNEITRAVGANGTPARGPQNNWAFYTGDD 89
50 KVFDAEIDN-----ANVSDSNKDGASYLIPQCKAEYKASTDFNSLFTTTNGRTYTVTK 104
-----
Db 90 YTIADVADHIIITAINVADTTPIGNI-----AQNTVWGSIVTGCNLLPVITAGKSLTLN 145
QY 105 KDTASANEIATWAKSISANTTPVSTVTSNNNDGTEVINVSQYGYVYSSTVNGGAVIMV- 163
Db 146 GNNADAANHGFGAP--ADNYTGLGNIALGAGNAALIIQSAAPAKITLAGNINGGIITVK 203
QY 164 TSVTPNATIEKNTDATWGDGGKTVQDKTYSVGDTVK-----YTIYKNAV----- 210
Db 204 TDAINGTIGNTNALATVNVGAGIATLEGAIKATTTKLTNAASVLTNTVNAVLTGAID 263
QY 211 NYHGTKEYQYVIKDTMPSASVVDLNEGSYEVTITDGSN-ITLTTOGSEKAT----- 262
Db 264 NTTGVQDNV-----GVNLNGALSOVTGNIGNTNALATISVGAGKATLGGAVIK 311
QY 263 -GKYNLLENNNFITIPWAAATNT--PTGNTONGANDFFYKGINITIT--VTYTGVLKS- 316
Db 312 ATTTKLTDNASAVTFNPVVVTGAIDNTGNANNGI---VTFTGDSVTGNTGNALATI 368
QY 317 -----GA--KPGSADLPENTNIATI-NP-----NTSNDPQKVTVRDGOIT 355
Db 369 SVGAGKATLGGAIKATTTKLTDNASAVTFNPVVVTGAIDNTGNANNGIIVTFTGDSVT 428
QY 356 -----IKKID-GSTRASIQGAIIVLKNATGQFLNFDNNVWEGTEANATEYT----- 402
Db 429 GNIGNTNALATISVGAGKATLGGAI--IKATTTKL-----TDN-----ASAVTFNPV 475
QY 403 -TGA-----DGIITITGLKE-----GTYVLVEKKAPLG-----YNLLD 434
Db 476 VTGADNTGNANNGIIVTFTGDSVTGNTGNALATISVGAGKATLGGAIKATTTKLT 535
QY 435 NSQKV-----ILGDGATDTT-NSDMLLVNPTVENNKGTLPSTGGIGTTTFIYGAILVI 488
Db 536 NASAVTFNPVVVTGAIDNTGNANNGIIVTFTGDS-----VTGNIGNTNAL-----ATISV 586
QY 489 GAG 491
Db 587 GAG 589
RESULT 7
EGT2_YEAST STANDARD; PRT; 1041 AA.
AC P42835;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE EGT2 protein precursor (Early G1 transcript 2).
GN EGT2 OR YNL327W OR N0320.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ON NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1676;
RX MEDLINE=95373280; PubMed=7645347;
RA Maftahi M., Nicaud J.-M., Levesque H., Gaillardin C.;
RT "Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV
RT identifies the RPD3, PAS8 and KRE1 loci, five new open reading
RT frames."
RL Yeast 11:567-572 (1995).
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=96251274; PubMed=8668141;
RA Kovacech B., Nasmyth K., Schuster T.;
RT "SGT2 gene transcription is induced predominantly by Swi5 in early
RT G1."
RL Mol. Cell. Biol. 16:3264-3274 (1996).
CC -1- FUNCTION: Seems to be involved in the correct timing of cell
CC separation after cytokinesis, as separation of mutant daughter
CC cells is delayed. Could either be an enzyme necessary for glucans-
CC degradation of the cell wall at the neck region between mother and
```

daughter cells or a regulatory protein controlling this metabolic step. Exclusively expressed between the end of mitosis and early G1; inactivated before cells pass start.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Z46259; CAA86371.1; --  
 EMBL; Z71603; CAA96259.1; --  
 PIR; S55862; S55862.  
 GenOnline; 143333; --  
 SGD; S0005271; EGT2.  
 GO; GO:009277; C:cell wall (sensu Fungi); IDA.  
 GO; GO:000910; P:cytokinesis; IMP.  
 GO; GO:000910; P:cytokinesis; IMP.  
 Glycoprotein; Repeat; Signal; Cell cycle; Cell division; Mitosis.

FT SIGNAL 1 20  
 FT CHAIN 21 1041  
 FT EGT2 PROTEIN.  
 FT DOMAIN 200 203  
 FT POLY-SER.  
 FT DOMAIN 381 384  
 FT POLY-SER.  
 FT DOMAIN 388 395  
 FT POLY-SER.  
 FT DOMAIN 490 493  
 FT POLY-SER.  
 FT DOMAIN 586 589  
 FT POLY-THR.  
 FT REPEAT 457 492  
 FT 1-1.  
 FT REPEAT 577 606  
 FT 1-2.  
 FT REPEAT 613 647  
 FT 1-3.  
 FT REPEAT 716 745  
 FT 1-4.  
 FT REPEAT 773 802  
 FT 1-5.  
 FT REPEAT 811 840  
 FT 1-6.  
 FT REPEAT 849 886  
 FT 1-7.  
 FT REPEAT 887 924  
 FT 1-8.  
 FT REPEAT 925 962  
 FT 1-9.  
 FT CARBOHYD 65 65  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 103 103  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 161 161  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 175 175  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 249 249  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 332 332  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 401 401  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 435 435  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 465 465  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 485 485  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 506 506  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 526 526  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 544 544  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 556 556  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 635 635  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 636 636  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 657 657  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 709 709  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 756 756  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SEQUENCE 1041 AA; 108494 MW; 01PECF8AA8744CD CRC64;

Query Match 6.9%; Score 178; DB 1; Length 1041;  
 Best Local Similarity 20.3%; Pred. No. 0.013;  
 Matches 104; Conservative 83; Mismatches 218; Indels 108; Gaps 16;

QY 24 TPIAFAAETGTTITVDTQKATKAY-KVFDIAIDNAN-----VSDSNKDGASLYLP 74  
 DB 437 TTTAVTQSTSTIAVSSAEKLSLTSTVSNVTSVSSATQHTTTPSVSNSTLTSSSVLE 496  
 QY 75 QGKEASYKASDPSNLSFTTTTNGGRVTVKOTASANEIATWA-----KSISA--N 123  
 DB 497 SVISSPYLANTTVSGASSASQSTNPPYVSNSTSSATQLATTAFAINTGTSSISSITN 556  
 QY 124 TTPVSTVTSNNDGTEVINVS-QGYVYVSVSTVNGAVIMVTSVTPNATHEKNTDATWG 182  
 DB 557 TSSVSSSTSSSGPPFVSNVAVSGSYILTTTESAQLTEIGSLIPSTI-----TSTTT 612

QY 183 DGGKTVQKTVSGDVTKYITYKNAVNYHTEKYQYVVIKDTWPSASVDLNEGSYEV 242  
 DB 613 SGTDKTGNKVAS-----STEAQSVNNSSLSVSTINTNAATAAA 653  
 QY 243 TITDGSNITLTQSGERATGKYNLENNFTIIPWAATNTPTGNTONGANDDDFFYKG 302  
 DB 654 NARNATFTHTSGSLQPSYSSLSLSS-----TIDTKVTATTSTSRDGSLSLAFTTG 707  
 QY 303 INTITVTYV-----GVLKSGAKPGSADLPENTNIATINPTNSNDPQKVTVRDQGITI 356  
 DB 708 LMQSVVTGDKSDTYSVISSTESAQVTEYDLSLPISTLKPTV----- 749  
 QY 357 KKIDGSTKASLQGAFLVKATQGLNFNDNTNNVWTE-ANATEY----- 401  
 DB 750 --VTGTSRNTSTFMSVSSTKLTEATATDKGDAYSIVSSQSAQVTEYGSMLPISTLETPTV 807  
 QY 402 --TTGADGIIITGLKEGYVLYLVEKAPLGVNLLNS-----OKVILG-----DGATDT 448  
 DB 808 IMSTDSEGYFLTTCETE-SGQATEYGLSLPISTLDGSLVYTFGESVVGYSVTVGAQY 866  
 QY 449 TNSDMLLVNPTVENNK---GTPLSTGGIGTTI 478  
 DB 867 AQTSLVPSTIKGSKTSLSTESVVGYSVTV 899  
 RESULT 8  
 CNA\_STAAU  
 ID CNA\_STAAU STANDARD; PRT; 1183 AA.  
 AC Q53654;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Collagen adhesin precursor.  
 GN CNA.  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FDA 574;  
 RX MEDLINE=92165839; PubMed=1311320;  
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,  
 Lindberg M., Hoeoek M.;  
 RT "Molecular characterization and expression of a gene encoding a  
 Staphylococcus aureus collagen adhesin.";  
 RL J. Biol. Chem. 267:4766-4772(1992).  
 RN [2]  
 RP ERRATUM.  
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,  
 Lindberg M., Hoeoek M.;  
 RL J. Biol. Chem. 269:11672-11672(1994).  
 RN [3]  
 RP COLLAGEN-BINDING DOMAIN.  
 RC STRAIN=FDA 574;  
 RX MEDLINE=94032261; PubMed=8218209;  
 RA Patti J.M., Boles J.O., Hoeoek M.;  
 RT "Identification and biochemical characterization of the ligand  
 binding domain of the collagen adhesin from Staphylococcus aureus.";  
 RL Biochemistry 32:11428-11435(1993).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.  
 RX MEDLINE=97475225; PubMed=9334749;  
 RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,  
 Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,  
 Narayana S.V.L.;  
 RT "Structure of the collagen-binding domain from a Staphylococcus  
 aureus adhesin.";  
 RL Nat. Struct. Biol. 4:833-838(1997).  
 CC -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO  
 COLLAGEN-CONTAINING SUBSTRATA.  
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 an amide bond (Potential).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----

DR EMBL; M81736; AAA20874.1; --  
DR PDB; 1AMX; 24-JUN-98.  
DR PDB; 1D20; 27-SEP-00.  
DR PDB; 1D2P; 27-SEP-00.  
DR InterPro; IPR008966; Adhes bact.  
DR InterPro; IPR008454; Cna\_B  
DR InterPro; IPR008970; Cna\_B unit.  
DR InterPro; IPR008456; Collagen bind.  
DR InterPro; IPR001899; Gram\_pos anchor.  
DR Pfam; PF05738; Cna\_B; 7.  
DR Pfam; PF05737; Collagen bind; 1.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; Repeat; Signal; 3D-structure.  
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.  
FT SIGNAL 1 29  
FT CHAIN 30 1154 COLLAGEN ADHESIN.  
FT PROPEP 1155 1183 REMOVED BY SORTASE (POTENTIAL).  
FT DOMAIN 151 318 COLLAGEN-BINDING.  
FT DOMAIN 533 1093 3 X 187 AA APPROXIMATE TANDEM REPEATS.  
FT DOMAIN 1093 1157 LYS/PRO-RICH (CELL WALL-SPANNING).  
FT REPEAT 533 719 B1.  
FT REPEAT 720 906 B2.  
FT REPEAT 907 1093 B3.  
FT SITE 1151 1155 LPXTG SORTING SIGNAL (POTENTIAL).  
FT MOD RES 1154 1154 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
FT STRAND 174 179  
FT TURN 182 183  
FT TURN 185 186  
FT STRAND 187 194  
FT TURN 196 197  
FT STRAND 201 201  
FT STRAND 205 211  
FT STRAND 215 228  
FT STRAND 229 230  
FT STRAND 232 234  
FT HELIX 239 246  
FT TURN 248 249  
FT STRAND 251 255  
FT TURN 256 259  
FT STRAND 260 265  
FT HELIX 267 270  
FT TURN 271 272  
FT STRAND 273 283  
FT TURN 286 287  
FT STRAND 290 299  
FT STRAND 301 301  
FT TURN 302 303  
FT STRAND 307 311  
FT STRAND 314 317  
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072B575D76 CRC64;

Query Match 6.9%; Score 177.5; DB 1; Length 1183;  
Best Local Similarity 21.8%; Pred. No. 0.017;  
Matches 152; Conservative 65; Mismatches 236; Indels 245; Gaps 35;  
QY 24 TPIAFAETGTTVDQTK--GATYKAYKVFDAEIDN-----AN----- 60  
DB 512 TPEGTKKENGVLVNTKEPTTTSISGEKWD-DKQNDGKRPEKVSVLLANGKVKVT 570  
QY 61 ---VSDSN-----KDGASYLPQKGEAYKASTFNSLFTTTTNGGRFYVTKDTAGANE 112  
DB 571 LDVTSSTWNVKFEKDLPKY--DEGKKIETVTEHDVVDYTTDING--TTITNKYTPGETS 626  
QY 113 IATWAKSISANTT-----PVSTVTESSNDG-----TEVINVS-----Q 145

DB 627 -ATVTKNWDNNQDGRKPTTEIKVELYQDGKATKTAALNESNNWTHWTGLDEKAKGQQ 685  
QY 146 XGYI-----YVSSTVNN--GAVIMVTSVTNATHEKNTDATWGD---GGGKTVDQ 191  
DB 686 VKYIVVELTKVGTYTHVDNDMGNLIVNKYTPETT--SISGEKVWDDKDNQDGKREK 743  
QY 192 -----KTVSGDVTVKYITTYKNAVNYHGTEKVOYVTKDTPSPASVVDLN--- 236  
DB 744 VSVNLLADGEKVKTLDTSETNWKYEFKDLPKYDEGKKI-EYTVTEHDVVDYTTDINGTT 802  
QY 237 -----EGSYEVYIT--DGSGN-----ITLITOGSEKATGYNLLENNNFTTIP 279  
DB 803 ITNKYTPGETSATVTKNWDNNQDGRKPTTEIKVELYQDGKATKTAALNESNNWTH-- 860  
QY 280 WAATNTPT-----GNTQNGANDDFFYKGINITITVYT----- 311  
DB 861 WTGLDEKAKGQOVKYTVBELTKVGTYTHVDNDM--GNLIVNKYTPETTSISGEKVM 917  
QY 312 -----GVLSGAKPGSADLPENTNIATINPTNSDDPGQV--TVRDGQ 353  
DB 918 DDKNQDGRPEKVSVLLANGKVKTLDTSETNWKYEFKDLPKYDEGKKIETVYTEDH 977  
QY 354 -----ITIKIDGSTKASLOGAIFVLKN-----A 377  
DB 978 VKDYTTDINGTTITNKYTPGETSAT-----VTKNWDNNQDGRKPTTEIKVELYQDGKA 1031  
QY 378 TQOFLNFNDYNN-----VEWGTENATEYTTGADG-----IIT- 410  
DB 1032 TGTATLINESNNWTHWTGLDEKAKGQOVKYTVDELTKVNG--YTHVDNDMGNLIVTN 1089  
QY 411 -ITGLKEGTYLYVEK---KAPLGYNLLDNSQKVLGDGATDTNSDNLVNPTEVNNKGT 466  
DB 1090 KYTPKPKNPPIYPEKPKDKTPTPKDHSNKVKVTPPKPKSVKDDQDQKDKNKPKENPLK 1149  
QY 467 ELPSTGG--IGTTFIVIIIGAILVIGAGIVLVARRRLRS 502  
DB 1150 ELPKTYGKMTIITSWITWFIGIL-----GLYLILRKXFN 1183  
RESULT 9  
PMP6\_CHLPN STANDARD; PRT; 1276 AA.  
AC Q92899; Q9JRW2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable outer membrane protein pmp6 precursor (Polymorphic membrane protein 6).  
GN PMP6 OR CPN0444 OR CP0309 OR CPB0460.  
OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia. phila.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CNL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [3]



```
FT REPEAT      506 577 E (TYPE II).
FT REPEAT      578 652 F (TYPE II).
FT REPEAT      653 724 G (TYPE II).
FT REPEAT      725 799 H (TYPE II).
FT REPEAT      800 874 I (TYPE II).
FT REPEAT      875 949 J (TYPE II).
FT REPEAT      950 1021 K (TYPE II).
FT REPEAT     1022 1093 L (TYPE II).
FT REPEAT     1094 1165 M (TYPE II).
FT REPEAT     1166 1180 TYPE I (INCOMPLETE).
SQ SEQUENCE    2249 AA; 224333 MW; A9D6464C089DF087 CRC64;

Query Match
Best Local Similarity 6.6%; Score 169; DB 1; Length 2249;
Matches 140; Conservative 53; Mismatches 190; Indels 164; Gaps 32;

Qy 12 ASLAFGMASVPTPIAPAAETGTTTQDTQKATYKAYKVFDAEIDNA----NVDSNKKD 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 741 ATLGGAIVKATTTKLTNAASVLT-----NANAVLTGAIDNTTGGDNVGVNLN 790

Qy 68 GASVLIPOGKEAEYKASTDF---NSLET-----TTTNGGRTVYTKDTASANEIATWAK 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 GAL-----SQVTGDIQNTSLATISVGAGTATLGA--VIKATTTKLTNAASVLT 838

Qy 119 SISAN---TTPVSTVTESNDGTEVINVSQVGYVYSSTVWNGAVIMVTSVTPN---AT 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 839 LTNAVLTGAVDNTTGGDNVGV---VLNL-----NGALSQVTGDIQNTSLAT 893

Qy 172 IHEKNTDATWGDGGKTVDDQKYSVGTVKYTIYKNAV-----NVHGTETKVVQYVVKD 225
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 884 ISVGAGTATLGGAVIKATTTKLTNAASVLT---LTNANAVLTGAIDNTTGGDNV----- 934

Qy 226 TMPASVVDINEGSEYETITDGSN--ITVLTQSEKAT-----CKYNLLENNNNFTI 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 935 -----GVNLNLGALSQVTGDIQNTSLATISVGAGTATLGAIVKATTTKLTDAASAVKF 989

Qy 277 TIPWAAINT--PTGNTQNGANDDFYKGINITITVYTVGLKSGAKPGSADLPENTN--IAT 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 990 TNPVVVTGAIDNTGNANG-----IVTFG---NSIVTGVNGV---NTNALAT 1030

Qy 334 INPNTSNDPGQKTVTRDQGITIKIDGSKYAS-----LOGAIFVLKNAITGQPLNF 384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1031 VNVGAG-----LLQVQGVVVKANTINLTNDNASAVTFTNPVVVTGAIDNTGNANGIVTF 1084

Qy 385 NDTNVEWGTETANATEYTT---GADGLIITIG---LKEGTYLVEKKAPLGVNLLDINSQV 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1085 TGNSTVT-GNVGNTNALATVNVGA-GLLQVQGVVVKANT-----INLTDNASAV 1131

Qy 440 -----ILGDGATDTT-NSDNLVV-----NPTVENNKGTELPSTGGIGTTFIYGAILVIG 489
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1132 TFTNPVVVTGAIDNTGNANGIVTFGTGNSIV-----TGDIQNTNAL---ATVNVG 1178

Qy 490 AGIVLVA 496
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1179 AGITLQA 1185

RESULT 11
EAE_ECO27
ID -EAE ECO27 STANDARD; PRT; 939 AA.
AC P19809;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Incimin (Attaching and effacing protein) (Eae protein).
GN EAE OR EAEA.
OS Escherichia coli O127:H6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_taxID=168807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O127:H6 / E2348/69;
```

```
RX MEDLINE=91045893; PubMed=2172966;
RA Jerse A.E., Yu J., Tall B.D., Kaper J.B.;
RT "A genetic locus of enteropathogenic Escherichia coli necessary for
RL the production of attaching and effacing lesions on tissue culture
RL cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7839-7843(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O127:H6 / E2348/69;
RX MEDLINE=98254123; PubMed=9593291;
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RT "The complete sequence of the locus of enterocyte effacement (LEE)
RT from enteropathogenic Escherichia coli E2348/69.";
RL Mol. Microbiol. 28:1-4(1998).
CC -|- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
CC LESIONS ON TISSUE CULTURE CELLS. BELIEVED TO MEDIATE ADHERENCE.
CC -|- SUBCELLULAR LOCATION: Outer surface.
CC -|- SIMILARITY: Belongs to the intimin/invasin family.
CC -|- SIMILARITY: Contains 1 LysM repeat.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; MS8154; AA62775.1; -.
DR EMBL; AF022236; AAC38392.1; -.
DR PIR; I41197; I41197.
DR PDB; 1ESU; O3-OCT-00.
DR PDB; 1F00; 12-JUL-00.
DR PDB; 1F02; 12-JUL-00.
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003343; Big_2.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR002482; LysM.
DR Pfam; PF02369; Big_1; 2.
DR Pfam; PF02368; Big_2; 1.
DR Pfam; PF00059; lectin c; 1.
DR Pfam; PF01476; LysM; 1.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID_1; 2.
DR SMART; SM00635; BID_2; 1.
DR SMART; SM00257; LysM; 1.
KW Outer membrane; Virulence; 3D-structure.
FT REPEAT 65 113 LYSM.
SQ SEQUENCE 939 AA; 102410 MW; 783C53EB0322DE4D CRC64;

Query Match 6.5%; Score 167.5; DB 1; Length 939;
Best Local Similarity 24.1%; Pred. No. 0.044;
Matches 116; Conservative 42; Mismatches 195; Indels 129; Gaps 23;

Qy 27 AFAAETGTTTVDQTKGATYKA-----YKVFDAEIDNANVSNSKDGASVLP 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 ALRSQGGQIQHSQSAQDYQAILPAYVGGSNVYKVTARAYDRNGNSNNVLTITVLS 550

Qy 75 QGKEAEYKASTDFNSLFTTTTNGGRTYVTKDTASANEIATWAKSISANT---TPVSTVT 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 NGQVQVQVGTDFADTKTSADGTEAITVATVKKGVAQANVPVSVFVSGTAVLSAN 610

Qy 132 ESNDG-----TEVINVSQYGYVYS-----STVNGAVIMVTSVTPNATHEKNTDAT 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 611 SANTNGSGKATVTLKDFQGVVVSAKTAEMTSALANAVIFVDQT--KASITEIKADKT 668

Qy 181 WGDGGKGTVDQKTSV---GD-----TVKYTIT---YKNVNVHGTETKVVQYVVKD 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 669 TAVANGQ--DAITYTVKVMKGDVPVSNQEVFTTTLGLKLSNSTEKTDITNGYAKVLTSTT 726
```

QY 228 PSASVVDLN-----EGSYEVITDGSNITTLTGSEKAT-----GKNLLE 269  
 Db 727 PGKLSARVSVDVADKAPVEFEFTLLTDGNIIEIVGTGVKGLPTVWLQYGVNLKA 786  
 QY 270 ENNNFTITPWAATNPTGNTQNGAND-DFFYKGINTI-----TVYITGVLKSGAKP 320  
 Db 787 SGGNGKVT--WRSANPAIASVDASSQGVTLKEGKTTTISVSSDNTQTAYT----- 835  
 QY 321 GSADLPENTNIATIN-----PNTSNDNDPGQKVTVRDGOITIKKIDGSKASLOGAIFVLKN 376  
 Db 836 -----IATPNSLIVPNNS-----KRVTYNDVAVNTCKNPGGKLPSSQ----- 871  
 QY 377 ATGQFLNFNDTNNV--EMGTEANATYTTGADGII-----TITGLKEG---TYVLVEKA 426  
 Db 872 -----NELENVFRWAGA-ANKYBYKSSQTIISWVQOQAQDAKSGVASTVDLV-KQN 921  
 QY 427 PL 428  
 Db 922 PL 923

RESULT 12  
 ID PIP LACLC STANDARD; PRT; 1902 AA.  
 AC P16271;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PI-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine  
 DE proteinase).  
 GN PRTP.  
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
 OG Plasmid pW05.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 OX NCBI\_TaxID=1359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WG2;  
 RX MEDLINE=88149035; PubMed=3278687;  
 RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;  
 RT "Nucleotide sequence of the cell wall proteinase gene of  
 RT Streptococcus cremoris Wg2";  
 RL Appl. Environ. Microbiol. 54:231-238 (1988).  
 CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
 CC GROWTH OF THE BACTERIA ON MILK.  
 CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
 CC specificity, although some substrate preference have been noted,  
 CC e.g. large hydrophobic residues in the P1 and P4 positions, and  
 CC Pro in the P2 position. Best known for its action on caseins,  
 CC although it has been shown to hydrolyze hemoglobin and oxidized  
 CC insulin B-chain.  
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 CC an amide bond (Potential).  
 CC -!- SIMILARITY: Belongs to peptidase family S8.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M24767; AAA17677.1; -;  
 CC HSSP; P00782; 1S01.  
 CC MEROPS; S08.019; -;  
 CC InterPro; IPR001899; Gram\_pos\_anchor.  
 CC InterPro; IPR003137; PA.  
 CC InterPro; IPR00209; Peptidase\_S8.  
 CC Pfam; PF00746; Gram\_pos\_anchor; 1.  
 CC Pfam; PF02225; PA; 1  
 CC Pfam; PF00082; Peptidase\_S8; 1.  
 CC PRINTS; PR00723; SUBTILISIN.

TIGRFBMS; TIGR01167; LPXTG\_anchor; 1.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;  
 KW Signal; Plasmid.  
 FT SIGNAL; 1 33  
 FT PROPEP 34 187  
 FT CHAIN 188 1870  
 FT PROPEP 1871 1902  
 FT ACT\_SITE 217 217  
 FT ACT\_SITE 281 281  
 FT ACT\_SITE 620 620  
 FT ACT\_SITE 1867 1871  
 FT SITE 1870 1870  
 FT MOD\_RES 1870 1870  
 SQ SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5DOB CRC64;  
 Query Match 6.5%; Score 167.5; DB 1; Length 1902;  
 Best Local Similarity 21.6%; Pred. No. 0.11;  
 Matches 114; Conservative 71; Mismatches 171; Indels 173; Gaps 26;  
 QY 25 PIAFAAETGTTTVDQTKGATYKAVKVFDAEIDNANVSDSNKDGASYLIPQCKEAEYKAS 84  
 Db 704 PSTVAENGYPAVELKDFSTDTKPKL-----TFTNSTTHELY-----QMDSN 747  
 QY 85 TDFNSLFTTTT--NGRTYVTKKDTASANEIATWAKSISANTTPVSTVTE----- 132  
 Db 748 TDTNAVYTSATDPNSGVLYDKKIDGAA-----IKAGSNITVPAGKTAIEFTLSLPKS 800  
 QY 133 -----SNNDGTEVINVSQYGY-----YVSTVNGAVIMV 163  
 Db 801 FQOQOFVEGFLNFKGSDGSR-LNLPYMGFFGDMNDGKIVDSLNGITYSPAGNGFGTVPL 859  
 QY 164 TGVTPNATHEKNTDATWGDG-----GGTKVDOKTYSVGDVTYKVTITVKNVNVHGTEK 217  
 Db 860 TN-----KVTGTQYGGVMTDADGNQTVDDQAIASFSD-----KNAL-YNDISM 902  
 QY 218 VYQYVVKDTPMSAVVDLNEGSEYVITDGSN-ITTLTGSEKATGKYNLLLENNFTI 276  
 Db 903 KY-YLLR-----NISNVQVDILDGQGNKVTLSSTNLTKTYNAHSQQYIYN 950  
 QY 277 TTPWAAT--NTPTGNTQNGANDDPYKGINTIYTVYTVGLKSGAKPGSADLPENTNIAT 334  
 Db 951 APAMDGTYYDQDQGNIKTADDSYTYR-----ISGVEGGDKRQVDFVPFKLD--- 998  
 QY 335 NENTSNDPQGVTVRVDGOITIKKIDGST-----KASLOG--AIFVLKNATGQFLNF 384  
 Db 999 -----SKAPTVRHVALSAKTENGKTQYVLTAEAKODLSGLDGTKSVKTA-----I 1043  
 QY 385 NDTNNVWNGTEANATEYTTGADGIITI-TGLKEGTYYLVKRAKPLGYNLLDNSQKVLGD 443  
 Db 1044 NEVTNL---DATFTDAGTTADGYTKIETPLSD-----EQAQALNG--DNSAELVLT 1091  
 QY 444 GATDTTNSDLNVNPTVENNKTELPSTGGIGTTFYIIGAILVIGAGI 492  
 Db 1092 NASNATDQASQVKP-----GSTSPDLI-----VNGGGI 1120

RESULT 13  
 YDBA\_ECOLI  
 ID YDBA\_ECOLI STANDARD; PRT; 2003 AA.  
 AC P33666; P76087; P76856; P76857; P76859;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein ydbA.  
 GN YDBA OR B1401/B1405.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RA "The complete genome sequence of Escherichia coli K-12.";  
RT Science 277:1453-1474 (1997).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
Kasai H., Kashimoto K., Kimura S., Kitagawa M., Kitagawa M.,  
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
Sampai G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RT DNA Res. 3:363-377 (1996).  
RL [3]  
RN SEQUENCE OF 464-2003 FROM N.A.  
RP STRAIN=K12;  
RX MEDLINE=92190338; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RT "Multiple IS insertion sequences near the replication terminus in  
Escherichia coli K-12.";  
RL Biochimie 73:1361-1374 (1991).  
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).  
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR  
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT  
BETWEEN AMINO ACIDS 839 AND 840.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE000237; AAC74483.1; ALT SEQ.  
DR EMBL; AE000237; AAC74487.1; ALT SEQ.  
DR EMBL; D90778; BAA15009.1; ALT SEQ.  
DR EMBL; D90778; BAA18880.1; ALT SEQ.  
DR EMBL; D90779; BAA18881.1; ALT SEQ.  
DR EMBL; X62680; -; NOT ANNOTATED CDS.  
DR EcoGene; EG11307; ydha.  
KW Hypothetical protein; Complete proteome.  
FT CONFLICT 489 489 I -> V (IN REF. 2).  
FT CONFLICT 495 495 I -> V (IN REF. 2).  
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C6B53220EE CRC64;  
Query Match 6.5%; Score 167; DB 1; Length 2003;  
Best Local Similarity 22.3%; Pred. No. 0.12;  
Matches 136; Conservative 67; Mismatches 196; Indels 210; Gaps 29;  
QY 33 GTTITVDTQKATYKAYKVFDAEIDANVSDSKDGSALYLPQCKEAYKASTDFNSLFT 92  
DB 279 GTMTVDPES-----MGIIQID-GDKAVNNEGSTINGG-----TGTQINGDDA 322  
QY 93 TTTNGGRVTVTKKDTASANEIATWAKSISANTTPVSTVTSNNNDGTEVI-----NVSOYG 147  
DB 323 TANNNGKITVDGKD-----STGTEINGNNGKVIQDGLDVSQGG 361  
QY 148 YYY-----VSSVNNAGVIMVTSVT-----PNATHEKNTDATWGGGKTVQDKTYS 195  
DB 362 HGIDITGSDVNTKGTMTVTDPEISIGIQVGDQAVNNEGSAITNGTGTQINGDDAT 421  
QY 196 VGDVTKYTYTKYKAVNVNHGTE-----KYYQ-----YVTKDTPMSASV----- 232

DB 422 ANNNGKTTVDGKOST---GTEIAGNNGKVIQDGLDVSQGGHGIDITGDSATVDNKGTMT 478  
QY 233 -----VDLNEGSYEVTITDG-----SGNITTLTQOSE----- 259  
DB 479 VTDPSIGIQIDGDOAIVNNEG--ESTITNGTGTQINGNDATANNNGKTTVDGKDSGTG 536  
QY 260 KATGKYNLENNPTIT-IPWAATNTPTGTQNGANDDFPKYKINTITVTVT---GVLK 315  
DB 537 KIAGNIGIVNLDGSLTVTGGAGVENIGDNGTVNNGKD-----IIVSDTGSIGVLI 587  
QY 316 SCAGKPSADLPENTNIA-TINPTNSNDPQKVTVRDGOITI---KKIDGSTRKASLQ- 369  
DB 588 NG-----EGATVSNITGDVNVNNEATGFSITNSKSVSLAGSMQVDFSTGVDLNGN 638  
QY 370 -----AIFVLK-----NATQQLNFNDTNV---EMGTEANATEY-----T 402  
DB 639 NNSVTLAAKDLKVVQKATGIVNSGDANTVITGNVLVDKDTADNAAEYFPDPSGVIN 698  
QY 403 TGADGIITITGLKEGTYYLVEKAPLGYNLLDNSOK-----VILGDGATDTTNSDLAV- 456  
DB 699 YGSDNNVTLDG--KLTVVSDSEVTSRQNLFDGSAEKTSGLVVIGDNTVNMNGGLELIG 756  
QY 457 --NPTVENNKGTLPSTGGTITFIYIIG-----AI 485  
DB 757 EKNALADGSQVTSL-RTGYSYTSVIVSGESSVYLNNGDTTISGEFPLGFAGVIRVQDKAL 815  
QY 486 LVIGAGIVL 494  
DB 816 LBIGSGATL 824  
RESULT 14  
P2P LACLC  
ID P2P LACLC STANDARD; PRT; 1902 AA.  
AC P15293;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-  
associated serine proteinase) (LPI511).  
DE associated serine proteinase)  
GN PRT.  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OG Plasmid pLP763  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=NCDO 763;  
RX MEDLINE=89313288; PubMed=2501630;  
RA Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashima A.;  
RT "Molecular characterization of a cell wall-associated proteinase gene  
from Streptococcus lactis NCDO763.";  
RL Mol. Microbiol. 3:359-369 (1989).  
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
GROWTH OF THE BACTERIA ON MILK.  
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
specificity, although some substrate preference have been noted,  
e.g. large hydrophobic residues in the P1 and P4 positions, and  
pro in the P2 position. Best known for its action on caseins,  
although it has been shown to hydrolyze hemoglobin and oxidized  
insulin B-chain.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (Potential).  
CC -!- SIMILARITY: Belongs to peptidase family S8.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----



Q53020; Q9ZCM0;  
30-MAY-2000 (Rel. 39, Created)  
30-MAY-2000 (Rel. 39, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Outer membrane protein B precursor (168 kDa surface-layer protein)  
(Surface protein antigen) (Cell surface antigen 5) (Scas) (rOMPb)  
(omp B) (Contains: 120 kDa surface-exposed protein (Surface protein  
antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).  
ONMP OR SPAP OR SPA OR RP704.  
Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=782;  
[1]  
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=Breil;  
RC MEDLINE=91045972; PubMed=2122457;  
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
RT "Characterization of the gene encoding the protective paracrystalline-  
RT surface-layer protein of Rickettsia prowazekii: presence of a  
RT truncated identical homolog in Rickettsia typhi";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
RN [2]  
SEQUENCE FROM N.A.  
RC STRAIN=Breil;  
RC Moron C.G., Yu X.J., Walker D.H.;  
RT "Sequence analysis of ompB of Rickettsia prowazekii.";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
SEQUENCE FROM N.A.  
RC STRAIN=Madrid E;  
RC MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RL mitochondria";  
RN Nature 396:133-140(1998).  
[4]  
PARTIAL SEQUENCE.  
RC STRAIN=Breil;  
RC MEDLINE=92114896; PubMed=1370573;  
RA Ching W.M., Carl M., Dasch G.A.;  
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of  
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
RT prowazekii";  
RL Mol. Immunol. 29:95-105(1992).  
[5]  
CLEAVAGE SITE.  
RC MEDLINE=92104668; PubMed=1729180;  
RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;  
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
RT membrane protein of rickettsiae: identification of an avirulent mutant  
RT deficient in processing";  
RL Infect. Immun. 60:159-165(1992).  
CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -1- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-  
CC layer with hexagonal symmetry.  
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR ENBL; M37647; AAA26390.1; ALT INIT.  
DR ENBL; AF161079; AAD42234.1; --



```
Query Match 6.4%; Score 165; DB 1; Length 1655;
Best Local Similarity 22.8%; Pred. No. 0.12;
Matches 134; Conservative 67; Mismatches 209; Indels 182; Gaps 29;

QY 30 AETGTTTVDQTKGATYKAYKVFDAEIDNANVSDSN-----KDGA-----SYLIP 74
Db 606 AQTLLNGKIGTVGANKTLGQENIGSSKTVLSDGVAINELVIGNGAVQFAHNTYLLIT 665
QY 75 QGKEAEYKASTDPN-----SLFTTTTNGGRT-----YVTKKOTPASANEIATWAKSTS 121
Db 666 RTTNAAGQGGKIIFNPVNNNTTLLATGNLGSATNPLAEINFGSKGAANDVTVLNVGKGVN 725
QY 122 ANTTPVSTVTS-----NNDGTEVINVS---QYGYVYSSTVNNCAVIMVTSVTPNAT 172
Db 726 LYATNITITDANVGSPIFAGGTSIVSGTVGGQGNKFNFTVALDNGTITVKFLG---NATF 782
QY 173 HEKNT---DATWGDGGKTVDOKTYSVG-----DTVKYTIYK-----207
Db 783 NGNTTIAANSTLQIGNYTADFVASADGTGIVEFVNTGPIVTVLNKQAAPVNAKQITVS 842
QY 208 -----NAVNYHGTGKTYQYVYIKOTMPSASVVDLNEGSYEVTITDGSNITFL-- 254
Db 843 GPGNVVINEIGNAGNYHGA--VTDITAFENSLGAVVFLPRG---IPFNDAGNRIPLTIK 897
QY 255 -TQSEKATGKYNLLENNFTIIPWA-----ANTPTGNTQN-----GANDDF 298
Db 898 STVGKNTATG-----FDVPSVILVGVDSIADGQVIGDQNNIVGLGLGSDNDI 945
QY 299 -----FYKGINTI---TVTVTG-----VLKSGAKPGSADLPENT-----NIAT 333
Db 946 IVNATLYAGITGINNQGTVLUSGIPNTPGTVYGLTGIGASKPKQVTFDFTDYNLGN 1005
QY 334 INPNTSNDPFGQKTVYR-----DQGITIKKID-----GSTKASL 367
Db 1006 LIATNATINDGVTVTTGGIAGFDGCKITLGSVNGNVRVFDGILSHSTSMIGTTKAN- 1064
QY 368 QGAIFVLKNAQTGFLNFDNNTNVEWGTENATYTTGADGIIITITGLKGGTYVLEKAP 427
Db 1065 NGTVTVYLGNAFVGNIGSDT-----FVASRVFTGSDG---GAGLOGNIYSQV---ID 1110
QY 428 LG-YNLDNSQKVLGDGATDITNSDNLVNPVTNNKGTLPSTGIGITTI 478
Db 1111 FGTYNLGINSNVILGGGTGTAINGKINLRN-TLTFASGT---STWGNNTSI 1158

RESULT 17
P3P_LACLC STANDARD; PRT; 1902 AA.
AC F15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PIII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-
DE associated serine proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK11;
RX MEDLINE=89340435; PubMed=2760036;
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:13579-13585(1989).
CC -I- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -I- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some subsite preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC Pro in the P2 position. Best known for its action on caseins,
CC
```

```
although it has been shown to hydrolyze hemoglobin and oxidized
insulin B-chain.
-I- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
an amide bond (Potential).
CC -I- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; J04962; AAA03533.1; ALT_SEQ.
DR HSSP; P00782; 2SBT.
DR MEROPS; S08.019; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003337; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRfam; TIGR01167; LPXVG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW Signal; Plasmid.
FT SIGNAL 1 33
FT PROPEP 34 187 PIII-TYPE PROTEINASE.
FT CHAIN 188 1870 REMOVED BY SORTASE (POTENTIAL).
FT PROPEP 1871 1902 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 LPXVG SORTING SIGNAL (POTENTIAL).
FT SITE 1867 1871 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT MOD_RES 1870 1870
SQ SEQUENCE 1902 AA; 200550 MW; 87CECBA9A345F9D3 CRC64;

Query Match 6.4%; Score 164.5; DB 1; Length 1902;
Best Local Similarity 20.9%; Pred. No. 0.16;
Matches 108; Conservative 76; Mismatches 184; Indels 149; Gaps 25;

QY 25 PIAFAAETGTTITVDQTKGATYKAYKVFDAEIDNANVSDSNKDGASYLLIPQGEAEYKAS 84
Db 704 PSTVVAENGYPAVELKDFSTDTKFKL-----TFTNRTTHELT-----QMSDN 747
QY 85 TDFNSLFTTTT---NGGRTYVTVTKKDTASANEIATWAKSISANTTPVSTVTE----- 132
Db 748 TDTNAVYTSATDPNSGVLYDKIDGAA-----IKAGSNITVPAGKTAIEFTLSLPKS 800
QY 133 -----SNNDGTEVINVSQYGYVYSSTVNNCAVIMVTSVTPNAT----- 171
Db 801 FDOQGFVEGLNFKGSDGSR-LNLPMGPF---GDWNDGKIVDSLNGITYSPAGNFGTV 856
QY 172 --IHEKNTDATWGDG-----CGKTVDOKTYSVGDTVKYTIYKNAVNVHGTGKTYQYVI 223
Db 857 PLLKNKNTGTYGGMVTDGKNTVDDQAIAPSSD-----KNAL-YNDISMKY-YLL 907
QY 224 KDTMPSASVVDLNEGSYEVTITDGSN-ITTLTQSEKATGKYNLLENNFTIIPWA 282
Db 908 R-----NISNVQVLDLGGQNKVTLSSSTNRKKTYYNAHSQYIYNAPAWDG 956
QY 283 T--NTPGTNGTNGANDDPFYKGINTIITVYTVGLKSGAKPGSADLPENTNIATNPNTSN 340
Db 957 TYDQDQGNIKTADGGSYTR-----ISGVPEGGDKRQVDFVPFKLD----- 998
QY 341 DPGQKTVYRDCQITIKKIDGSTK-----ASLQGAFLVKNATGQFLNFDNTNVEWGTGA 396
Db 999 ---SKAPTVRHVALSAKTENGKTYVLTAAEKDDLSGLDATSKSVKTEINEVTNL-----DA 1051
```

QY 397 NATEYTTGAGIITI-TGLKEGYLYLVEKKAPLGYNLLNSQVILGDGATDTNSDNL 455  
 ID YFJA\_ECOLI STANDARD; PRT; 1569 AA.  
 AC P52143; P76610; P77017; P77019;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical outer membrane protein YpJA.  
 GN YFJA OR B2647.  
 QY 456 VNPTVENNKTELPSTGGIGTTFYIGAILVIGAGI 492  
 DB 1104 QKP-----GSTSFDLI-----VNGGGI 1120

## RESULT 18

YFJA\_ECOLI  
 AC P52143; P76610; P77017; P77019;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical outer membrane protein YpJA.  
 GN YFJA OR B2647.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R. Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RA "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [2]

## SEQUENCE FROM N.A.

RC STRAIN=K12;  
 RX MEDLINE=97349980; PubMed=9205837;  
 RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horiuchi T.;  
 RA "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 K-12 genome corresponding to 50.0-68.8 min on the linkage map and  
 analysis of its sequence features";  
 RL DNA Res. 4:91-113(1997).  
 CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).  
 CC -!- SIMILARITY: STRONG, TO BORDETTELLA PERTACTIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC -----  
 DR EMBL; U36840; AAA79815.1; ALT\_SEQ.  
 DR EMBL; AB000350; AAC75695.1; -.  
 DR EMBL; D90889; BAA16514.1; ALT\_INIT.  
 DR EMBL; D90890; BAA16518.1; ALT\_INIT.  
 DR PIR; A65044; A65044.  
 DR EcoGene; EG13213; ypJA.  
 DR InterPro; IPR006315; Autotransport.  
 DR InterPro; IPR005546; Autotransporter.  
 DR InterPro; IPR004899; Pertactin.  
 DR InterPro; IPR003991; Pertactin C.  
 DR Pfam; PF03797; Autotransporter; 1.  
 DR Pfam; PF03212; Pertactin; 1.  
 DR PRINTS; PR01484; PR01484.FAMILY.  
 DR TIGRPFAMs; TIGR01414; Autotrans barl; 2.  
 KW Hypothetical protein; Outer membrane; Complete proteome.

SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;  
 Query Match 6.4%; Score 164; DB 1; Length 1569;  
 Best Local Similarity 25.3%; Pred. No. 0.13;  
 Matches 137; Conservative 55; Mismatches 208; Indels 142; Gaps 30;  
 QY 32 TGTITV---QDTQXG--ATYKAYKVFDAIDNANVSDSNKDGASVLI-----PQKEAE 80  
 DB 159 TGT-TIESGNQDVYKGGISNGTTIKGASRVGGSGANGILIDGSGQIVKVGQHADGTTIN 217  
 QY 81 YKASTDF--NSLFTTTT-NGRTYV---TKKDTASANEIATWAKSISANTTTPVSTVTESN 134  
 DB 218 KSGSQDVVQGSLATNTTNGGROYEQSVETTTTKNGGEQRVYESRALDTTIEGQTSL 277  
 QY 135 NDGTEVINVSQY--GYVYVSSV-----VNGGAVIMVTSVTPNATIEKNTDATTWGDGG 185  
 DB 278 NKSSTAKNTHIVSGGTQIVDNITSTSDVIEVSGGVLDRGGTA-----TNVQHDS- 328  
 QY 186 GKTVPQKTVSGDVTVKYITTYKNAVNYHGTKEKYQYVIKDT-MPSASVVDLNE-GSYEVT 243  
 DB 329 ---AILKTNNGTTVSGT-----NSEGAFSIHNHVDNVLLENGCHLDINAYGSANKT 378  
 QY 244 IYDGSNITTLTQSGEKATGKYNLLENNNFTTIPWAATNT-PTCNTONGANDDFFYKG 302  
 DB 379 IIRKDKGMSVLTKAKADAT-----RIDNGGYMDVAGNATNTIINGGTQNNIN-----YG 427  
 QY 303 INTITVTVYTG--LKSGAK-----PGSADLPENTNIATINPNTS----- 339  
 DB 428 IATGTNNGSTQNIKSGGKADTTIISGSRQVVEKDGTA-IGSNISAGSLIVYTGIAH 486  
 QY 340 --NDDPQKQVTVRDGQITIKKIDGSTKAS---LOG--AIFVLKNATGQF----- 381  
 DB 487 GVNQETGSALVANTGAGT--DIEGYNKLSHFTITGGEANYVVLNTGELTVVAKTSAKNT 544  
 QY 382 -----LNFNNTNNVVEGTEANATVTTGADGIIITG--LKEGT- 418  
 DB 545 TIDTGKLIQKEAKTDSTRLLNNGGVLEVDQGEAKHVEQSGGALIASTTSGTLLEGTN 604  
 QY 419 -----YYLVEKKAPLGYNLLDN--SQKVLGDCATDTTNSDNLVN-----PTVENNKG 465  
 DB 605 SYGDAPYIRNSEAK--NVLENAGSLTVVTGSRVAVDTIINANGKMDVYKDVGTVLNSAG 662  
 QY 466 TE 467  
 DB 663 TQ 664

## RESULT 19

BIGAL\_SALTY STANDARD; PRT; 1953 AA.  
 ID BIGAL\_SALTY  
 AC P25927; P25928; Q9XCQ3;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative surface-exposed virulence protein bigA precursor.  
 GN BIGA OR STM3478.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 14028;  
 RA Stojiljkovic I., Valentine P., Heffron F.;  
 RT "Salmonella typhimurium rhs homolog.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,







RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror J.-J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzener T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*,"  
 RL Nature 390:249-256(1997).  
 CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,  
 CC MOTILITY, SECRETION OR DIFFERENTIATION.  
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED  
 CC INTO THE MEDIUM.  
 CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE  
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE  
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED  
 CC MOTIF REPEATED 31 TIMES.  
 CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME  
 CC SIMILARITY TO THE REPEAT IN E. COLI RHS GROUP OF PROTEINS (RHS-A-D).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L05634; AAA22893.1; -;  
 DR EMBL; D31856; BAA06656.1; -;  
 DR EMBL; D29985; BAA06260.1; -;  
 DR EMBL; D83026; BAA11683.1; -;  
 DR EMBL; Z99124; CAB15959.1; -;  
 DR PIR; S32920; S32920.  
 DR Subtilast; BG10797; wapa.  
 DR InterPro; IPR003305; CBM\_CenC.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF02018; CBM 4.9; 1.  
 DR Pfam; PF05593; RHS repeat; 14.  
 DR TIGRfam; TIGR01643; YD repeat 2x; 17.  
 KW Cell wall; Repeat; Signal; Complete proteome.  
 FT SIGNAL 1 28  
 FT CHAIN 29 2334  
 FT DOMAIN 504 869  
 FT REPEAT 504 605  
 FT REPEAT 636 736  
 FT REPEAT 769 869  
 FT DOMAIN 1021 2139  
 FT REPEAT 1021 1040  
 FT REPEAT 1042 1061  
 FT REPEAT 1063 1082  
 FT REPEAT 1083 1102  
 FT REPEAT 1109 1128  
 FT REPEAT 1129 1148  
 FT REPEAT 1150 1169  
 FT REPEAT 1174 1193  
 FT REPEAT 1199 1218  
 FT REPEAT 1219 1238  
 FT REPEAT 1646 1665  
 FT REPEAT 1667 1686  
 FT REPEAT 1690 1709  
 FT REPEAT 1711 1730  
 FT REPEAT 1732 1751  
 FT REPEAT 1753 1772  
 FT REPEAT 1795 1814  
 FT REPEAT 1820 1839  
 FT REPEAT 1840 1859  
 FT REPEAT 1861 1880  
 FT REPEAT 1887 1906

FT REPEAT 1908 1927 2-22.  
 FT REPEAT 1929 1948 2-23.  
 FT REPEAT 1969 1982 2-24 (APPROXIMATE).  
 FT REPEAT 1983 2002 2-25.  
 FT REPEAT 2008 2027 2-26.  
 FT REPEAT 2028 2047 2-27.  
 FT REPEAT 2051 2070 2-28.  
 FT REPEAT 2071 2090 2-29.  
 FT REPEAT 2093 2112 2-30.  
 FT REPEAT 2120 2139 2-31.  
 SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;  
 Query Match 6.2%; Score 159; DB 1; Length 2334;  
 Best Local Similarity 22.4%; Pred.No. 0.4; Indels 150; Gaps 30;  
 Matches 121; Conservative 67; Mismatches 201;  
 QY 32 TGTITVQDTQKATYKAVKVFDAEIDNANVSDNSKDGASYLIPQGEAEYKASTDFNSLF 91  
 DB 1593 TGTWVFD-----IRLEGLSLTKSTYDSN---GNYVTKEDELGYATSTDYDE-- 1638  
 QY 92 TTTTNGRTYVTKDPTASANEIATWAKSISANTTPVSTVTESNNDGTEVINVSQGYVYV 151  
 DB 1639 ---TGKKT---SETDAKGEKTVTYD---QADQLTNMTLSN---GTSILHSYDKEGNEV 1685  
 QY 152 SSTVANGA-----VIMVTSVTPN---ATIHENKNTDATWGDGGGKTVDQK---- 192  
 DB 1686 SKTIRAGADQTYKFEYDVNGKLVKTTDPLGNVLASEYDANSNLT-----KTISPNGNEV 1739  
 QY 193 --TYSVGDITVKYITTYKNAVNYHGKVVQYVIKDTMPSASVVDLNEGSEYEVITTDGSGN 250  
 DB 1740 SLISYDGTDRVK-----SKSYNGTEK-YIFYDKNGNETSVVNKEQNTTKRTFDNKR 1791  
 QY 251 ITTLT-----QGSKATGK---YNLENNNF-TITIPWAAVN 284  
 DB 1792 LTELDRGSGQTWTPSDSKLTPSWIHGDKGTNQFTYKLDQMIEMKDTSSYSFDY 1851  
 QY 285 TPTGNTQ-----NGANDDFYKGIINTVITYTGVLSKG-AKPGSADLPENTHAIATNPNT 338  
 DB 1852 DENGNVQTFITNGGGTFSYDERNLVSLHIGDKNGGDLTESYEYDANGNRTTINSSA 1911  
 QY 339 SND-----KVTVRDGOI-----TIKKIDGSTKASLQGAFLVK 375  
 DB 1912 SGKQVEYKLNQLVKETHEDGTVEIYDVGNGKRTVTIK--DGSKT--VNASF-- 1964  
 QY 376 NATGQFLNFDNTNNVWGTENATYTTGADGIITITGLKEGYLVLEK-----APLG 429  
 DB 1965 NIMNQLTKVND-ESISYDKNGNRT-----SDCKFTVTWDAEDNLTAVTKKGDKPFTYK 2018  
 QY 430 YNLLDNS-QKVILG-----DCATDTNSDNLNVPVNNKGTGLPSTGGIGTIFY 480  
 DB 2019 YDEKGNRIQKTVNGKVTNYFYDG-----DSLNLVYETDADNNV-TKSYTYGDSGQLLSY 2071  
 RESULT 22  
 SLPO\_BACBR STANDARD; PRT; 1004 AA.  
 AC P09333;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Outer cell wall protein precursor (OWP).  
 OS *Bacillus brevis* (*Brevibacillus brevis*).  
 OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; *Brevibacillus*.  
 OX NCBI\_TaxID=1393;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.  
 RC STRAIN=47;  
 RX MEDLINE=87008404; PubMed=2428810;  
 RA Tsuboi A., Uchihi R., Tabata R., Takahashi Y., Hashiba H., Sasaki T.,  
 RA Yamagata H., Tsukagoshi N., Uda S.;  
 RT "Characterization of the genes coding for two major cell wall  
 RT proteins from protein-producing *Bacillus brevis* 47: complete  
 RT nucleotide sequence of the outer wall protein gene.";



RL J. Bacteriol. 168:365-373(1986).  
RN [2]  
RP SEQUENCE OF 1-85 FROM N.A.  
RC STRAIN=47;  
RX MEDLINE=88115203; PubMed=2828336;  
RA Tauboi A., Uchih R., Adachi T., Sasaki T., Hayakawa S., Yamagata H.,  
RA Takagoshi N., Uda S.;  
RT "Characterization of the genes for the hexagonally arranged surface  
RT layer proteins in protein-producing Bacillus brevis 47: complete  
RT nucleotide sequence of the middle wall protein gene."  
RL J. Bacteriol. 170:935-945(1988).  
CC -I- FUNCTION: THE OUTER WALL PROTEIN BINDS TO THE MIDDLE CELL WALL  
CC PROTEIN.  
CC  
CC -I- SUBUNIT: THE OUTER CELL WALL LAYER IS COMPOSED OF SUBUNITS OF  
CC THE OUTER CELL WALL PROTEIN. THESE PROTEINS FORM A HEXAGONAL  
CC ARRAY WITH A LATTICE CONSTANT OF 14.5 NM IN THE OUTER CELL  
CC WALL LAYERS.  
CC  
CC -I- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-  
CC layer with hexagonal symmetry.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; M14238; AAA22373.1; -  
DR EMBL; M19115; AAA22761.1; -  
DR PIR; B25039; B25039.  
KW Cell wall; S-layer; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 1004 OUTER CELL WALL PROTEIN  
FT SEQUENCE 1004 AA; 106137 MW; 3534F65ED0EB9E CRC64;  
SQ  
  
Query Match 6.1%; Score 158; DB 1; Length 1004;  
Best Local Similarity 20.8%; Pred. No. 0.16;  
Matches 130; Conservative 75; Mismatches 202; Indels 218; Gaps 27;  
  
Qy 16 FGMAVSPVPIFAAETGT---ITVDTQKATYKAYKVFDAIDNANVSDSNKDGASYL 72  
Db 412 FAADTPPTTKVBAKNTNTEIHVTFSETVAGADNKANFTLKGVGN-----V 458  
Qy 73 IPQCKEAEYKASTDFNSLFTT-TTNGKRTVTKK--DTASANEIATWAKSIS-ANTTPVS 128  
Db 459 IPLTKAEVDAAKNYKVVTEPLNGGSYLVTKGIEDASKNKLVEYTAIVAVADIVPPN 518  
Qy 129 -----TVTESNN 135  
Db 519 VKLDLPATPGTDAQLISPTKVIAPTEPMDKASIKNNYMFNGFNLDKSVTLTATDSNT 578  
Qy 136 ----DGEVINVSQYGYYSSTVNGAVIMVTSVTPNATTHEKNTDWTGDSGGKTVQD 191  
Db 579 AVVVDFTNVVGFNGF-----KNGDAISVGRVLDTA-----GNPKTEMQ 616  
Qy 192 KTVSGDVTVKYTIYKNAVNYHGTKEYV-QYVIDTTPSASVVDLNBGSYE----- 241  
Db 617 TKVNLPSVSAFLPDKAEVTKNIVKLYFKELIINAKADDFAVDNGEG-YKAVNSISNDV 675  
Qy 242 -----VTIDSGNITLTGSEKATGKYNLLNENNNFTIIPWATNTP----- 286  
Db 676 VENKSVITLTGNDLPTTAAGVKVKTGGE---VDKKNQYGVAV---ALTDVPADDKIGPNW 730  
Qy 287 ----TGNTONGANDFFYKGINITITVTVTGVLKSGAKPGSADLPENTNIAINPTNSDD 342  
Db 731 LKAETVDTNNKIDQF-----KLTFSEALYVASVDSPFRIEGYTIAGV-----ET 777  
Qy 343 PQQKTVTRDGGQITIKKIDGSKASIQGAIFVLK-NATQGF-----LNFND-----TN 388  
Db 778 KGEVTVIKVTELDIDSDATPTVAIGSVGVDLKRNASGFFEPQKALDGVSAAPDKAPVWT 837  
Qy 389 NVEWTEANATEYTTGADGIIITGLKE-----GTYILV----- 422

Db 838 GVEAGKTYNTAVTPDSADKDIKTVLKKDQKELAGYALKTPISNGSYELVVTNAGNNT 897  
Qy 423 -----EKKAPLGYNLLDSQKVLGDG-----ATDTN---SDNLLVNPT 459  
Db 898 TVKFVKVDIPADSKKAPKTKTVD--KVAVADAPKWEAPKATATDDVDGDISDKIATVYS 955  
Qy 460 VEN--NKGTELPST---GGIGTTI 478  
Db 956 SEDAGSKVYDLASATHLGTAGNTV 980  
  
RESULT 23  
AG43 ECOLI  
ID AG43 ECOLI STANDARD; PRT: 1039 AA.  
AC P39180; P75614; P76360; P97241; Q46771;  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Antigen 43 precursor (AG43) (Fluffing protein).  
GN FLU OR B2000.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251358; PubMed=9097040;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,  
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,  
RA Nishimoto K., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,  
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,  
RA Yamamoto Y., Horiiuchi T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 40.1-50.0 min region on the linkage map."  
RL DNA Res. 3:379-392(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ML 308-225;  
RX Henderson I.R., Owen P.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP PRELIMINARY SEQUENCE OF 53-78.  
RC STRAIN=ML 308-225;  
RX MEDLINE=89291704; PubMed=2661530;  
RA Caffrey P., Owen P.;  
RT "Purification and N-terminal sequence of the alpha subunit of antigen  
RT 43, a unique protein complex associated with the outer membrane of  
RT Escherichia coli."  
RL J. Bacteriol. 171:3634-3640(1989).  
RN [5]  
RP SEQUENCE OF 53-63.  
RC STRAIN=K12 / EMG2;  
RX Link A.J., Robison K., Church G.M.;  
RT "Comparing the predicted and observed properties of proteins encoded  
RT in the genome of Escherichia coli K-12."  
RL Electrophoresis 18:1259-1313(1997).  
RN [6]  
RP GENE NAME.  
RX MEDLINE=97257509; PubMed=9103983;



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch).

CC EMBL; M75136; AAA88153.1; --  
DR PIR; F36791; F36791.  
KW Hypothetical protein; Repeat.  
FT REPEAT 143 158  
FT REPEAT 171 186  
FT REPEAT 200 214  
FT REPEAT 215 233  
FT REPEAT 234 252  
FT REPEAT 253 268  
FT REPEAT 279 293  
FT REPEAT 294 309  
FT REPEAT 320 334  
FT REPEAT 335 349  
FT REPEAT 362 376  
FT REPEAT 377 391  
FT REPEAT 392 406  
FT REPEAT 407 421  
FT REPEAT 422 436  
FT REPEAT 437 452  
FT REPEAT 464 477  
FT REPEAT 478 493  
FT REPEAT 504 517  
FT REPEAT 518 531  
FT REPEAT 532 545  
FT REPEAT 546 559  
FT REPEAT 560 573  
FT REPEAT 574 587  
FT REPEAT 588 601  
FT REPEAT 602 615  
FT REPEAT 616 629  
FT REPEAT 629 643  
SQ SEQUENCE 670 AA; 64174 MW; 2B64A781C51958B4 CRC64;

Query Match 6.1%; Score 157.5; DB 1; Length 670;  
Best Local Similarity 23.0%; Pred. No. 0.1;  
Matches 121; Conservative 46; Mismatches 225; Indels 133; Gaps 26;

QY 21 SPVTPIAAETGTTIVQDTOKGATYKAYKVPDAEIDNANVSDSNKDGASYLIPQKKEA 80  
DB 141 TPATP-AGANDPANITTA-TPAGAN-----DTANITTTATPAGAN----- 177

QY 81 YKASTDFNSLFTTTTNGGRTYVTKKDTASANEIAT--WAKSISANTTPVSTVTSNNDGT 138  
DB 178 -----DTANIITATPAGA-----NDTAVTTTSATPAGANDTAVTTTP-ATPAGANDTAN 225

QY 139 EVINVSQYGVYVSVTVNNGAVIMVTSVTPNATTHEKNTDATWGDGGKTVQKTSYVG- 197  
DB 226 DTVTVTTPAMPAGANDTANGTAVTTTPAMPAGANDTANITTTATPAGANDTAN-VTMPAGA 284

QY 198 -DTVKYT-----ITVKNAVNVHGTEKYQYVVIKDTMPSASVVDLINEGSEYVT 243  
DB 285 TDTVTVTTPAMPTGANDTANIITATPAGANDTANV-----TWPA-----GATDTV 329

QY 244 ITD-----GSGNITTLTGSEKATGKNYLLENNNFITTPWATNT-----PTGN 289  
DB 330 VTTTPAMPAGANDTANVT-----KPAGSTDTVTVT--TPAMPTGATDTVTVTTPAMPTGA 382

QY 290 TQ-----NGANDDFY-----KGINTITVTYGLKSGAKPGSADLPENTNIAT 333  
DB 383 TDTVTVTTPAMPPTGATDTVTVTTPAKPAGANGTVTTTPAMPAGA-----NDTVVT 433

QY 334 INPNT---SNDDPGQKVTVRDQGITIKKIDGTSKASLOGAIFVLKNATGQFLNFNDNNV 390  
DB 434 TAPATPAGANDT--ANVTKPTGATDTVTVTATVKPT--GATGTVTTTAKPTGANDTANV 489

QY 391 -----EWGTEANATEYTTGADGIIITGLK-----SGTYLVVEKKAPLGYN--LLDNSOKV 439  
DB 490 TRKPTGATGTVTTTAKPTGATGTVTVTAKTGPTGATGTVTTTAK-PTGANGTVTTTAKP 548

QY 440 ILDGGATDTTNSDNLVNPVTNNKGTLPSTGGIGTITFYIGA 484

CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
EMBL; M75136; AAA88153.1; --					
DR	PIR; F36791; F36791.				
KW	Hypothetical protein; Repeat.				
FT	REPEAT	143	158	1.	
FT	REPEAT	171	186	2.	
FT	REPEAT	200	214	3.	
FT	REPEAT	215	233	4.	
FT	REPEAT	234	252	5.	
FT	REPEAT	253	268	6.	
FT	REPEAT	279	293	7.	
FT	REPEAT	294	309	8.	
FT	REPEAT	320	334	9.	
FT	REPEAT	335	349	10.	
FT	REPEAT	362	376	11.	
FT	REPEAT	377	391	12.	
FT	REPEAT	392	406	13.	
FT	REPEAT	407	421	14.	
FT	REPEAT	422	436	15.	
FT	REPEAT	437	452	16.	
FT	REPEAT	464	477	17.	
FT	REPEAT	478	493	18.	
FT	REPEAT	504	517	19.	
FT	REPEAT	518	531	20.	
FT	REPEAT	532	545	21.	
FT	REPEAT	546	559	22.	
FT	REPEAT	560	573	23.	
FT	REPEAT	574	587	24.	
FT	REPEAT	588	601	25.	
FT	REPEAT	602	615	26.	
FT	REPEAT	616	629	27.	
SQ	SEQUENCE	670	AA; 64174	NW; 2B64A781C519B8B4	CRC64;
	Query Match		6.1%;	Score 157.5;	DB 1; Length 670;
	Best Local Similarity		23.0%;	Pred. No. 0.1;	225; Indels 133; Gaps 26;
	Matches	121;	Conservative	46;	Mismatches 121; Indels 133; Gaps 26;
QY	21	SPVTPFAFAETGTTIVQDTOKGATYKAYKVPDAEIDNANVSDSNKDGASYLIPQKKEAE	80		
DB	141	TPATP-AGANDTANITTA-TPAGAN-----DTANITATPAGAN-----	177		
QY	81	YKASTDFNSLFTTTNGRTYVTKDTSASANIAT--WAKSISANTTPVSTVTSNNDGT	138		
DB	178	-----DTANITATPAGA-----NDTAVTTTTSATPAGANDTAVTTTP-ATPAGANDTAN	225		
QY	139	EVINVSQYGVYVSVTVNNGAVIMVTSVTPNAT IHEKNTDATWGDGGKTVQDKTYSVG-	197		
DB	226	DTVVTTTPAMPAGANDTANGTAVTTTPAMPAGANDTANITATPTGANDTAN-VTMPAGA	284		
QY	198	-DTVKYT-----ITYKNAVNYHGTGVYQVYVTKDTPSASVVDLNEGSEYVT	243		
DB	285	TDIVTTTTPAMPPTGANDTANITATPAGANDTANV-----TMPA-----GATDV	329		
QY	244	ITD-----GSGNITLTQSEKATGKNLLBENNFTITIPWAATNT-----PTGN	289		
DB	330	VTTTPAMPAGANDTANVT---KPAGSTDTVVTT---TPAMPPTGATDVTTTTPAMPPTGA	382		
QY	290	TQ-----NGANDDFY-----KGINTITVTVTYGVLKSGAKPGSADLPENTNINAT	333		
DB	383	TDIVTTTTPAMPPTGATDVTTTTPAKPAGANGTVTTTTPAMPAGA-----NDTVVT	433		
QY	334	INPTN-----SNDDPGQKVTVRDGQITIKKIDGTSKASLOGAIFVLKNATGQFLNFNDNNV	390		
DB	434	TAPATPAGANDT--ANVTKEPTGATDVTTTATVTKPT--GATGTVTTTITAKPTGANDTANV	489		
QY	391	-----ENGTEANATEYTTGADGIITITGLK-----EGTYLVVEKKAPLGVN--LLDSQKV	439		
DB	490	TKPTGATGTVTTTITAKPTGATGTVTTTATVTKPTGATGTVTTTITAK-PTGANGTVTTTITAKP	548		
QY	440	ILGDGATDTTNSDNLNVTNPTVENNKGTELPSTGGIGTTTIFYIIGA	484		





Db 137 KKEPRQYPSFAESFNDNAYVLNRTSFGNGYYAGTWKSNKTSYTDACLT-----GR- 190  
QY 101 YTKKDTASA--NEIATWAKSISANTPPVS-----TVTESNNDGTEVINVSQYGYVV 151  
Db 191 YATPCYAGKLNIIITY--GLTKYDTTPASGNAGGVITGNGGTG-----NTSNG----- 240  
QY 152 SSTVANGAVIMVTSVTPNATIEHKNTDATMGDGGKTVDOKTSYSGDVTVKYITIKYNAVN 211  
Db 241 STSGSGG-----SATTTGTYTVKSGDSVMG-----ISHSGITMAQLIEWNNIKN 287  
QY 212 ---YHTEKVKYQVINKTWPSSVVDLNEGSYEVTITDGSNGNITLTQGE--KATG-KY- 265  
Db 288 NFIIYPG-----QKLTIKGGQ--SAGSSTTNGN--NASSGNTSGNTSSTGQATGAKYT 339  
QY 266 -----NLLEN---NNF-----TTIPWAAVTPTGNT- 290  
Db 340 VKSGDSVMKIANDHGISMNQLIEWNNIKNFYVPGQQLVSKGSSASGTSSTSTGNTS 399  
QY 291 ONGANDDFYKGINITVTVGLKSGAKPGSAD---LPENTNIATINPNTSNDDPGOK 346  
Db 400 SNTAN-----TGSTTSGSTYT--VRAGESVWSVSKFGISMNQLIOWNNIKNFYVPGOK 452  
QY 347 VTVRDGQITIKKIDGSTKASLOCAIFVLKNATGQFLNFNDTNNVWEGTEANAT---EYTT 403  
Db 453 LIVKGG-----SSSSNASTSTA-----NNKVTASNTSSSTATGQATYTV 491  
QY 404 GADGIITITGLKEGYVYLVEKAPGLYNLLDNS-----QKVLGSGDGTNTSNDLLVNP 458  
Db 492 KAGESVMGVANKNG-----ISMNQLIEWNNIKNFYVPGOKLIVKGGSSKA--SATATIKP 545  
QY 459 TVENNKGIETPGTGIGITFIYI 481  
Db 546 TASTPAST--TPASSTGDTKYTV 567

## RESULT 29

BCA\_STRAG STANDARD; PRT; 1020 AA.  
AC Q02192;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE C protein alpha-antigen precursor.  
GN BCA.  
OS Streptococcus agalactiae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1311;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A909 / Serotype Ia;  
RX MEDLINE=93066179; PubMed=1438195;  
RA Michel J.L., Madoff L.C., Olson K., Kling D.E., Kasper D.L.,  
RA Ausubel F.M.;  
RT "Large, identical, tandem repeating units in the C protein alpha  
antigen gene, bca, of group B streptococci";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:10060-10064(1992).  
CC -!- FUNCTION: MAY PLAY A ROLE IN BOTH VIRULENCE AND IMMUNITY.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
CC an amide bond (Potential).  
CC -!- MISCELLANEOUS: IDENTICAL REPEATING UNITS DEFINE PROTECTIVE  
CC EPITOPES AND MAY PLAY A ROLE IN GENERATING PHENOTYPIC AND  
CC GENOTYPIC DIVERSITY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M97256; AAA26848.1; -.

DR InterPro; IPR005877; Gpos\_Ysirk.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF04650; Ysirk\_signal; 1.  
DR PRINTS; PR00015; GPOSANCHOR.  
DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
DR TIGRFAMs; TIGR01168; Ysirk\_signal; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
KW Antigen; Cell wall; Peptidoglycan-anchor; Repeat; Signal.  
FT SIGNAL 1 41  
FT CHAIN 42 990 C PROTEIN ALPHA-ANTIGEN.  
FT PROPEP 991 1020 REMOVED BY SORTASE (POTENTIAL).  
FT DOMAIN 227 364 9 X 82 AA TANDEM REPEATS.  
FT REPEAT 965 976 INCOMPLETE.  
FT SITE 987 991 LPXTG SORTING SIGNAL (POTENTIAL).  
FT MOD\_RES 990 990 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
SQ SEQUENCE 1020 AA; 108705 MW; 78035594FA190D5B CRC64;  
Query Match 6.1%; Score 156.5; DB 1; Length 1020;  
Best Local Similarity 21.8%; Pred. No. 0.2; Indels 151; Gaps 26;  
Matches 126; Conservative 64; Mismatches 236;  
QY 24 TPIAFAAETGTTITVODTOKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQKAEYKA 83  
Db 492 TPVSDKEITDLVKIPDGSKV-----PTVVGDRPDTNVPDGHKVTVEVTPDGTDTVEV 546  
QY 84 ST-----DFNSLFTTTTNGGRTYVTKKDTASANEIATWAKSISANTPPVSTVT----- 131  
Db 547 TVHVTVPKVPDKKYDPT---GGTTVPQGTVPVSDKEITDLVK-IPDGSKVPTVVGDPR 602  
QY 132 ESNNDGTEVINVSQYGYVVSSTVANGAVIMVTSVTPNATIEHKNTDATMGDGGKTVDO 191  
Db 603 DTNVEPDGHKVTVE---VTYPDGT---KDTVEVTVHVTVPKVPDKKYDPT---GGTTVPQ 654  
QY 192 KT---YSVGDTVKYITIKNA-----VNYHGTGKVKYQVINKTWPSSVVDLNEG 238  
Db 655 GTPVSDKEITDLVKIPDGSKVPTVVGDPRDTNVPDGHKV---TVEVTP-----DGTKD 706  
QY 239 SYEVTI-----TDGSGNITLTQGEKATGYNLLENNNFTITIPWAAATNP 286  
Db 707 TVEVTVHVTVPKVPDKKYDPTGTTVPQGTVPVSDKEITDL-----VKIPDGSKVPT 759  
QY 287 TGNTONGANDDFYKGINITVTVGLKSGAKPGSADLPENTNIATINPNTSND--DP- 343  
Db 760 ---TVVGDRPDTNVPDGHKVTVEVT-----YPDGTGDTVEVTVHVTVPKVPDKKYDPT 810  
QY 344 GOKVTVRDG-----QIT--IKKIDGS-----TKA 365  
Db 811 GGETTVPGTVPVSDKEITDLVKIPDGSKVPTVVGDPRDPTNVPDGHKVTVEVTPDGTKD 870  
QY 366 SLQGAIFVLKNATGQFLNFDNT---NNVWEGTEANATEYT-----TGADGIITITGLKE 416  
Db 871 TVEVTVHVTVPKVPDKKYDPTGTTVPQGTVPVSDKEITDLVKIPDGSKVPTVVGDPR 930  
QY 417 GTYYLVEKAPGLYNLLDNSQKVLGSGDGTNTN-----SDNLLVNPTVE-----N 462  
Db 931 DTNVEPDGHKVTV-----EVTYPDGTGDTVEVTVHVTVPKVPDKKYDPTGKAQVN 981  
QY 463 NKGTELPSTGGIGTTFVIIGAILVIGAVILVARRR 499  
Db 982 GKGKLPATGENATPFFNVAULTIISVGLLSVSKKK 1018

## RESULT 30

APU\_THETU

ID\_APU\_THETU STANDARD; PRT; 1861 AA.

AC P38536;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Amylopullulanase precursor (Alpha-amylase/pullulanase) (Pullulanase

DE type 1) [includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan

DE glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 10:19:36 ; Search time 43 Seconds  
(without alignments)  
3683.492 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKMIQSLVSLVAFGMVAV.....GAILVIGAGIVLVARRLRS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

1: SPREMBL.25.\*

2: sp\_archaea.\*

3: sp\_bacteria.\*

4: sp\_fungi.\*

5: sp\_human.\*

6: sp\_invertebrate.\*

7: sp\_mammal.\*

8: sp\_mmc.\*

9: sp\_organelle.\*

10: sp\_phase.\*

11: sp\_plant.\*

12: sp\_rodent.\*

13: sp\_virus.\*

14: sp\_vertebrate.\*

15: sp\_unclassified.\*

16: sp\_rvirus.\*

17: sp\_bacteriap.\*

18: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2578	100.0	502	2 Q84A41	Q84A41 streptococ
2	399	15.5	525	16 Q86G6	Q86G6 bifidobacte
3	331	12.8	625	16 Q83L8	Q83L8 enterococcu
4	324	12.6	522	16 Q8XP10	Q8XP10 clostridium
5	279.5	10.8	705	16 Q8DYR6	Q8DYR6 streptococ
6	265.5	10.3	535	2 Q8212	Q8212 actinomycet
7	265.5	10.3	665	16 Q87SC2	Q87SC2 streptococ
8	261	10.1	533	2 Q9X4D9	Q9X4D9 actinomycet
9	259	10.0	533	2 Q9X4D8	Q9X4D8 actinomycet
10	258.5	10.0	614	16 Q9CHH3	Q9CHH3 lactococcus
11	255	9.9	553	16 Q81D71	Q81D71 bacillus ce
12	251	9.7	507	2 Q48707	Q48707 lactobacill
13	250	9.7	533	2 Q9X4D2	Q9X4D2 actinomycet
14	250	9.7	705	16 Q88UC5	Q88UC5 lactobacill
15	244.5	9.5	554	16 Q85E4	Q85E4 streptococ
16	244.5	9.5	554	16 Q850S9	Q850S9 streptococ

17	243	9.4	535	2 Q9X4D1	Q9X4D1 actinomycet
18	232.5	9.0	674	16 Q8E4C4	Q8E4C4 streptococ
19	232	9.0	793	16 Q8YAG7	Q8YAG7 listeria mo
20	219	8.5	1612	16 Q8Y591	Q8Y591 listeria mo
21	217	8.4	562	16 Q8Y5H7	Q8Y5H7 listeria mo
22	216	8.4	1806	16 Q829J3	Q829J3 listeria in
23	215.5	8.4	3441	16 Q89PB9	Q89PB9 bradyrhizob
24	214.5	8.3	681	16 Q92FA4	Q92FA4 listeria in
25	212	8.2	2551	16 Q8CY18	Q8CY18 streptococ
26	208.5	8.1	476	16 Q836L9	Q836L9 enterococcu
27	208	8.1	917	16 Q88UJ0	Q88UJ0 lactobacill
28	206	8.0	520	16 Q8FMP6	Q8FMP6 corynebacte
29	205	8.0	1622	16 Q929J4	Q929J4 listeria in
30	204	7.9	3346	16 Q7WN54	Q7WN54 bordetella
31	200.5	7.8	827	16 Q8CPL9	Q8CPL9 staphylococ
32	200.5	7.8	893	16 Q97SC3	Q97SC3 streptococ
33	200	7.8	1441	16 Q9CFL1	Q9CFL1 lactococcus
34	199	7.7	1530	16 Q8Y479	Q8Y479 listeria mo
35	197.5	7.7	1006	16 Q88ZV1	Q88ZV1 lactobacill
36	195	7.6	2215	16 Q7WBN0	Q7WBN0 bordetella
37	194	7.5	1778	16 Q8FCB2	Q8FCB2 escherichia
38	193	7.5	454	16 Q8G453	Q8G453 bifidobacte
39	192.5	7.5	571	16 Q8YAG6	Q8YAG6 listeria mo
40	192	7.4	1744	16 Q82YW8	Q82YW8 enterococcu
41	191	7.4	1711	16 Q8Y6L8	Q8Y6L8 listeria mo
42	191	7.4	2000	16 Q814R4	Q814R4 bacillus ce
43	189.5	7.4	5017	16 Q81FU0	Q81FU0 clostridium
44	189	7.3	723	16 Q8XP11	Q8XP11 clostridium
45	189	7.3	1519	2 Q48237	Q48237 helicobacte
46	188.5	7.3	628	2 Q9FD50	Q9FD50 streptococ
47	187	7.3	2402	2 Q9AER7	Q9AER7 staphylococ
48	186.5	7.2	1093	16 Q81ANG	Q81ANG bacillus ce
49	186.5	7.2	1275	5 Q76602	Q76602 caenorhabdi
50	186.5	7.2	1733	2 Q9X114	Q9X114 staphylococ
51	186	7.2	1036	2 Q8RM89	Q8RM89 streptococ
52	186	7.2	3705	16 Q8ZHA1	Q8ZHA1 versinia pe
53	186	7.2	3710	16 Q8CZJ2	Q8CZJ2 synecococc
54	186	7.2	10791	16 Q7UW8	Q7UW8 synecococc
55	185.5	7.2	583	2 Q45822	Q45822 clostridium
56	185.5	7.2	1610	16 Q8FKP3	Q8FKP3 escherichia
57	185.5	7.2	1633	16 Q8CMP4	Q8CMP4 staphylococ
58	185	7.2	2768	16 Q8E9G6	Q8E9G6 shewanella
59	184.5	7.2	1330	2 Q934I7	Q934I7 microscilla
60	184.5	7.2	2468	16 Q912M3	Q912M3 pseudomonas
61	183.5	7.1	1183	16 Q8NUH0	Q8NUH0 staphylococ
62	182	7.1	775	2 Q9X6M6	Q9X6M6 enterococcu
63	182	7.1	865	2 Q9FDJ1	Q9FDJ1 streptococ
64	181.5	7.0	1315	2 Q85488	Q85488 staphylococ
65	181.5	7.0	1371	16 Q8XQ42	Q8XQ42 ralstonia s
66	181	7.0	1129	16 Q8XL11	Q8XL11 clostridium
67	181	7.0	3029	16 Q55582	Q55582 synecocyst
68	181	7.0	5010	16 Q81CV2	Q81CV2 bacillus ce
69	180.5	7.0	3930	16 Q98E20	Q98E20 rhizobium l
70	180	7.0	1166	2 Q86489	Q86489 staphylococ
71	180	7.0	1231	2 P72362	P72362 streptococ
72	180	7.0	1389	16 Q8E1C4	Q8E1C4 streptococ
73	180	7.0	2353	2 P71401	P71401 haemophilus
74	180	7.0	5017	16 Q81SN0	Q81SN0 bacillus an
75	179.5	7.0	1260	2 Q9XDB6	Q9XDB6 streptococ
76	179	6.9	1141	16 Q8NXX5	Q8NXX5 staphylococ
77	179	6.9	1307	16 Q8EV70	Q8EV70 mycoplasma
78	179	6.9	4106	16 Q8XQP2	Q8XQP2 ralstonia s
79	178.5	6.9	1614	2 Q9KKB6	Q9KKB6 astrakhan r
80	178.5	6.9	2566	17 Q8TSE7	Q8TSE7 methanosaarc
81	178.5	6.9	4936	16 Q8YKJ3	Q8YKJ3 anabaena sp
82	178	6.9	1195	3 Q96WU8	Q96WU8 schizosacch
83	178	6.9	3705	2 Q9F285	Q9F285 versinia pe
84	177.5	6.9	1268	2 Q9RER7	Q9RER7 bacillus sp
85	177.5	6.9	1965	16 P73089	P73089 synecocyst
86	177	6.9	440	2 Q07120	Q07120 lactobacill
87	177	6.9	1029	2 Q52708	Q52708 rickettsia
88	177	6.9	1347	16 Q8NXX6	Q8NXX6 staphylococ
89	176.5	6.8	1385	16 Q99W47	Q99W47 staphylococ

90	176.5	6.8	1794	16	Q8RHH1	Q8rhh1 fusobacteri												
ALIGNMENTS																		
RESULT 1																		
Q84A41																		
ID	Q84A41	PRELIMINARY;	PRT;	502 AA.														
AC	Q84A41																	
DT	01-JUN-2003 (TrEMBLrel. 24, Created)																	
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)																	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)																	
DE	SPB1.																	
DE	Streptococcus agalactiae.																	
GN	SPB1.																	
OS	Streptococcus agalactiae.																	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;																	
OC	Streptococcus.																	
OX	NCBI_TaxID=1311;																	
RN	[1]																	
RC	SEQUENCE FROM N.A.																	
RC	STRAIN=874391;																	
RA	Adderson E.E., Bohnsack J.F.;																	
RT	"Subtractive Hybridization Identifies a Novel Adhesin/Invasin of																	
RT	Virulent Serotype III Group B Streptococcus agalactiae."																	
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.																	
DR	EMBL; AF485279; AAC04909.1; -																	
DR	InterPro; IPR001899; Gram_pos_anchor.																	
DR	Pfam; PF00746; Gram_pos_anchor; 1.																	
DR	TIGRFAMS; TIGR01167; LPXTG_anchor; 1.																	
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.																	
SQ	SEQUENCE 502 AA; 53450 MW; E3B7F5603F6D6192 CRC64;																	
Query Match 100.0%; Score 2578; DB 2; Length 502;																		
Best Local Similarity 100.0%; Pred. No. 2.5e-109;																		
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;																		
Qy	1	MKKMIQSLVASLAFGMAVSPVTPIAFAAETGTTIVDTQKGATYKAYKVFDAEIDNAN	60															
Db	1	MKKMIQSLVASLAFGMAVSPVTPIAFAAETGTTIVDTQKGATYKAYKVFDAEIDNAN	60															
Qy	61	VSDSNKDGASYLIPQGEAEYKASTDFNSLFTTTTNGGRTYVTKOTASANEIATWAKSI	120															
Db	61	VSDSNKDGASYLIPQGEAEYKASTDFNSLFTTTTNGGRTYVTKOTASANEIATWAKSI	120															
Qy	121	SANTTPVSTVTESNNDGTEVINVSQGYVYVSVTVNNGAVIMVTSVTPNATIEHKNTDAT	180															
Db	121	SANTTPVSTVTESNNDGTEVINVSQGYVYVSVTVNNGAVIMVTSVTPNATIEHKNTDAT	180															
Qy	181	WGDGGKTVDPQKTVSGDVTVKYITITYKNAVNYHGTEKYQYVVKDTMPSASVVDLNEGSY	240															
Db	181	WGDGGKTVDPQKTVSGDVTVKYITITYKNAVNYHGTEKYQYVVKDTMPSASVVDLNEGSY	240															
Qy	241	EVITDGSNITTLTQGEKATGKYNLLENNFTIIPWAATNPTGNTONGANDDDFFY	300															
Db	241	EVITDGSNITTLTQGEKATGKYNLLENNFTIIPWAATNPTGNTONGANDDDFFY	300															
Qy	301	KGINTIVTVTVGLKSGAKPGSADLPENTNIATINPTNSNDPQKTVVRDGOITIKKID	360															
Db	301	KGINTIVTVTVGLKSGAKPGSADLPENTNIATINPTNSNDPQKTVVRDGOITIKKID	360															
Qy	361	GSTKASLQGAIFVLKNATGFLNFDNNTNNVEWGTTEANATEYTTGADGIITITGLKEGTY	420															
Db	361	GSTKASLQGAIFVLKNATGFLNFDNNTNNVEWGTTEANATEYTTGADGIITITGLKEGTY	420															
Qy	421	LVEKAPLGNLNDNSOKVILGDGATDTTNSDNLNVPNTVENNKGTELPTGGIGTIFY	480															
Db	421	LVEKAPLGNLNDNSOKVILGDGATDTTNSDNLNVPNTVENNKGTELPTGGIGTIFY	480															
Qy	481	IIGAILVICAGIVLVARRRLS 502																
Db	481	IIGAILVICAGIVLVARRRLS 502																
RESULT 2																		
Q8G6G6																		
ID	Q8G6G6	PRELIMINARY;	PRT;	525 AA.														
AC	Q8G6G6																	
DT	01-MAR-2003 (TrEMBLrel. 23, Created)																	
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)																	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)																	
DE	SPB1.																	
DE	Possible cell surface protein similar to FimA fimbrial subunit of																	
DE	Actinomyces naeslundii.																	
GN	BL0675.																	
OS	Bifidobacterium longum.																	
OC	Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;																	
OC	Bifidobacteriaceae; Bifidobacterium.																	
OX	NCBI_TaxID=216816;																	
RN	[1]																	
RP	SEQUENCE FROM N.A.																	
RC	STRAIN=NCC 2705;																	
RC	MEDLINE=22294977; PubMed=12381787;																	
RA	Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,																	
RA	Pessi G., Zwhilen M.-C., Desiere F., Bork P., Delley M.,																	
RA	Pridmore R.D., Arigoni F.;																	
RT	"The genome sequence of Bifidobacterium longum reflects its adaptation																	
RT	to the human gastrointestinal tract."																	
RL	Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).																	
DR	EMBL; AE014689; AAN24496.1; -																	
DR	InterPro; IPR008454; Cna_B.																	
DR	InterPro; IPR001899; Gram_pos_anchor.																	
DR	Pfam; PF00746; Gram_pos_anchor; 1.																	
DR	TIGRFAMS; TIGR01167; LPXTG_anchor; 1.																	
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.																	
KW	Complete proteome.																	
SQ	SEQUENCE 525 AA; 55559 MW; 11D954D398C55F75 CRC64;																	
Query Match 15.5%; Score 399; DB 16; Length 525;																		
Best Local Similarity 27.3%; Pred. No. 1.1e-10;																		
Matches 160; Conservative 72; Mismatches 198; Indels 156; Gaps 26;																		
Qy	3	KKMIQSLVASLAFGMAVSPVTPIAFAAETGTTIVDTQ---KGATYKAYKVFDAEIDNA	59															
Db	6	KVAAAGLAAATMGLGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	65															
Qy	60	NVSDSNKDGASYLIPQGEAEYKASTDFNSLFTTTTNGGRTYVTKOTASANEIATWAKSI	109															
Db	66	---GQATYTLASPDWDFKNADELGITGVTDAVSEKAYDYVYSL	109															
Qy	110	---ANEIATWAKSIANTTPVSTVTESNNDGTEVINVSQY-----GYYY	150															
Db	110	NQDMKLABFATKASNAWAKGAGNVSANPAATAKASG-----SQYVATFANLPLGYV	163															
Qy	151	VS---STVN---NGAVIMVTSVTPNATIH-----EKNDA-TWGDG---CGKTVDOK	192															
Db	164	VSPEAGSTNKRHTDAMLVNVTATTNINLKSEYPTVDTIDADKKGDSAQIGSKVNFQL	223															
Qy	193	TVSVGTVTKYITITYKNAVNYHGTEKYQYVVKDTMPSASVVDLNEGSYEVTITDGSNIT	252															
Db	224	KSKVPTSEY-----NYVFKIVDTLSAG--LDENN---DVTVKVGDAITLT	264															
Qy	253	TLTQGEKATGKYNLLENNFTIIPWAATNPTGNTONGANDDDFFYKGINITVITYTG	312															
Db	265	ATTDSVTTTKG-----TVTI-----DLSNVYKTDNASKAGKILVYSA	304															
Qy	313	VLKSGAKPGSADLPENTNIATIH-----NPNTSNDPQKTVVRDGOIT---IKKI-DGS	362															
Db	305	TLNENAFVGTDPQNNPGLNLSAKVQYVNGFSEENIGESTPSETHSYTFNFKKIYKGD	364															
Qy	363	TKASLQGAIFVLKNATGFLNFDNNTNNVEWGTTEANATEYTTGADGIITITGLKEGTY	409															
Db	365	TENALAGAKFQLLDSKTVISLVKKSNDVYRPAKTSDDTEV-----TEVETPATGII	416															
Qy	410	TITGLKEGTYVLEKAPLGNLNDNSOKVILGDGATDTTNS-DNLVN---PT-----	459															

```
Db 417 EFTGLKAGTYLLKETPAPKYNKLSDPVKVTINATINKTTGALESWTNVSAPTADVTVP 476
QY 460 ---VENNKTELPSTGGIGTITFIYIGAILV-IGAGIVLVARRR 501
Db 477 VVKIENKKGALLPDTGGMGTVLFTFGVLIVAGAGWYVKSXR 522

RESULT 3
Q83618
ID Q83618 PRELIMINARY; PRT; 625 AA.
AC Q83618;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Cell wall surface anchor family protein.
GN EF1093.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AB016950; AAO80893.1; -.
DR TIGR; EF1093; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGSFAM; TIGR01167; LPXG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Complete proteome.
SQ SEQUENCE 625 AA; 67989 MW; EB97450BB11F0DE CRC64;

Query Match 12.8%; Score 331; DB 16; Length 625;
Best Local Similarity 26.1%; Pred. No. 1.6e-027;
Matches 148; Conservative 78; Mismatches 227; Indels 114; Gaps 26;

QY 3 KKMIOQLLVASLAFGNVSPVPIAFAA--ETGTTTVDQTKGATYKAYKVFDAEIDNAN 60
Db 100 KQAVQSL-----TPGKPVAGCTDANGNTVOLPKQNGKDAVYTIKEPKEGV 148

QY 61 VDSNKGDA--SYLIPQKGEABYKASTD-----FNSLFTTTTNG--G 98
Db 149 VAATNMVAFPPVYEMIKOTDGSYKYGTEBLAVVHIYPKNVANDGSLKKVGTAEENGLNG 208

QY 99 RYVTKDPTASANEIA-----TWA-----KSISANTTPV--STVTSNNDGTE 139
Db 209 AEFVLSKSESGSGTKYIOGVKGLYTTWTDKEQAKRFTIGKSYSGENDFTEAEN--GTG 267

QY 140 VTNVS--QYGYTVVS--STVNNGAVIMVTSVTP-----NATIEHK--NTDATGPGGKG 187
Db 268 ELTVKNLEGVSVLIEVKAPNAELIENQTKTPFTIEANNQTPVEKTVKNDTSKVDKTP 327

QY 188 TVDQKTSYSGDVTVKYTYTYKNVAVNYHGTB----KVYQVYIKDTPMSASVVD--LNBSGYEV 242
Db 328 SLDGKDVAIKEIKYQISVNIPLGIADKGDANKYKFNVLVDKHAALTDFDNVTSGEYAY 387

QY 243 TITDGSNITLTITQSEKATGKYNLENNNTIYI--PWAATNPTGTNGANDDPFY 300
Db 388 ALYDGDVTI-----APENYQVTEQANGFTVAVNPAYIPTLTTPGG----- 426

QY 301 KGINTITVTYTVGLSKAGPGSADLPENTNIATINPNTSDPDGQKTVTRDQGITIKKID 360
Db 427 ----TLKFVYFHLNEKADPTKG-----FKNEANVDNGHTDDQPTPTVEVTGGRPIKVD 478
```

```
QY 361 GSTKA--SLOGAIFVLKNA---TGOFLNPDNTNN-VEW-GTEANATEYTTGADGIITNG 413
Db 479 GCVTATQALAGASFVVRQNSDTANYLKIDETTKAATWVKAEATFTTTADGLVDITG 538
QY 414 LKEGTYVLYVEKKAPLGYNLLDNSQKVLIGDGTATDTNSDNLVNPVTEN-NKGTELPSTG 472
Db 539 LKYGTYYLEETVAPDDYVLLTNRIEFVNVNEQSGYGT--ENLVSPEKVPKHKGT-LPSTG 595

QY 473 GIGTTFIYIIGAILVIGAGIVLVARR 499
Db 596 GKGIYVILGSGAVLLIAGVVFARRK 622

RESULT 4
Q8XP10
ID Q8XP10 PRELIMINARY; PRT; 522 AA.
AC Q8XP10;
DT 01-WAR-2002 (TReMBLrel. 20, Created)
DT 01-WAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Probable surface protein.
GN CPE0156.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79862.1; -.
DR InterPro; IPR008454; Cna_B.
DR Pfam; PF05738; Cna_B; 1.
KW Complete proteome.
SQ SEQUENCE 522 AA; 57247 MW; 8076613B81C314EA CRC64;

Query Match 12.6%; Score 324; DB 16; Length 522;
Best Local Similarity 25.7%; Pred. No. 2.7e-017;
Matches 143; Conservative 66; Mismatches 187; Indels 160; Gaps 27;

QY 19 AVSPVPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNANVSDSNKDGASVLIPOGKE 78
Db 32 SISKDAPI-----KGSIT--SKKGATFTAYKLDA-----IKSGDA 66

QY 79 ABYKASTDPNSLFTTTTNG--GRYVTKDPTASANEIATWAKSISANTTPVSTVTSNND 136
Db 67 YEYSVNSDLKDFNNSNYGSYSQESIQKLNGEQVKEFA-----INLH----KVILENKS 117

QY 137 GTEV-----INVSQYGYVYVSSTVNN--GAVIMVT-----SVTP--NA 170
Db 118 GOELKDGQKNTVDLGYLVTETSDSDEGAASVASTPIIVSPQVSGDSMNYDVTINPKDNT 177

QY 171 TIHEKNVDATWGDGGKTVQKTSYSGDVTVKYTYT-----YKNAVNYHGTGTEKVYQYV 224
Db 178 PILEKNIVKE-----NQRVKTSSSENIQDVVKVEKASIPVYQKNAQN-----IMYKFT 225

QY 225 DTMPASVVDVLENGSEYVYITDG--SGNITLTITQSEKATGKYNLENNNTFTIPWAA 283
Db 256 DTMSKGLTYDEKGT--FKVTSQDKVPAKDTDYTDVVK-----QEDGSTVITI----- 271

QY 284 NPTGNTONGANDDPFYKGINV-----ITVYTVGLVSKAGKPGSADLPENTNIA-----TI 334
Db 272 -----NFVYENIKAYAEYTGITLNYQATLNKDAVISNKENLGNLTNNIQLDYN 318

QY 335 NPNT--SNDPQKQTVTRDQGITIKKIDGSKAS--LOGNAIFVLKNATGOFL---NFNDIN 388
Db 319 NPHVKDSYKULTDKVTTTTFGTGFTKVDSELSNKLQGAEFVSKDAGGKIVAKYTYDEKG 378
```

```
QY 389 NVEWTEANATEYTTGADGIITITGLKEGTYLVEKKAFLGVNLLDNSOKVIL----- 441
Db 379 QVYV-LSNGG---VTNSKGIITFLGLKEGKYFTIEEVAPSGYLLKNPVEVITANKDES 434
QY 442 -----GDGATDTTNSD-NLLVNPVTVENKKGTELPSTGGIGTTTPIYIG 483
Db 435 GNYTGAATIEISNGKAGQIINDISEKDGNIILFNVOIENHAGFSLPSTGGLNGTGFIKIA 494
QY 484 AIL-----VIGAG 491
Db 495 IILLSIVCVLAIIUGL 510

RESULT 5
Q8DYR6 PRELIMINARY; PRT; 705 AA.
ID Q8DYR6
AC Q8DYR6;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Cell wall surface anchor family protein.
GN SAG1407.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Masiognani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Allen S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014256; AAN00278.1; -
DR TIGR; SAG1407; -
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMS; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Complete proteome.
SQ SEQUENCE 705 AA; 75475 MW; 7BB1774ECBBE1A1A CRC64;

Query Match 10.8%; Score 279.5; DB 16; Length 705;
Best Local Similarity 24.3%; Pred. No. 3.9e-05;
Matches 149; Conservative 76; Mismatches 213; Indels 175; Gaps 30;

QY 19 AVSPVTP-----TAFRAE--TGITVODTKGATY-----KAYKVFDAEIDANV 61
Db 132 AVSLTPEGTGLVNTKGLKGEFKIVEKSKSYNNNGSLLAASKAVPV-----NITL 195
QY 62 SDSNKGDA---SYLIPOGKEA-----EYKASTDFNSLFTTTTNGRTYVTKD 106
Db 186 PLVNEGDGVADAHVYPKNTKEPKIDKNFAKTNLTALTVDNRLTAGANYG-NYARDKA 244
QY 107 TAGAN-----ELAT-----WAKSISANTPVTSTVESNNDTEV----- 140
Db 245 TATAEIGKVVPYEVKTKIHKRSKYENLVWTDIMSNGLTMSGTVSLKASGTTTFAXD 304
QY 141 -INVSQYGY-----YVSSVTNNGAVIMVTSVTENATHEK 175
Db 305 ELSIDARGFTLKFTADGLGKLEAKAATADIEFLTYSATVNGQAIL-----DN 356
```

```
QY 176 NTDATWGGGGKTKVDQ--KTVSVGDTVKYITTYKNVNVHGTQYQYVIKD----- 225
Db 357 DIKLSYGNKPGKDELTPVPESKGE-VTVAKTWSGDIAPDGVNVVYTLKDKDKTVA 415
QY 226 TWPSASVVDLNEG-SYEVITIDSGSNITTLTG-----SEKATGKNLLE-EN 275
Db 416 TKTSKGTIDLNGIKFEV-----SGNFGSKFTGLENKSMYISERVSGYSGA 470
QY 276 ITIPWAATN-TPTGNTQ-----NGANDDFYKGINITITVITYTG---VLK 318
Db 471 INTKDSNDPTPLNTEPEKVTGHKKFKVKTNEQGBRL--AGAQFVVKN 528
QY 319 KPGSADLPE-----NTNIATINPNTSNDPPQKQVTVRDGQIITIKKIDG 373
Db 529 SEGQKTLAAKIALDEAIAAAYNKLGSATDQKGEK-----GITAKELIK 582
QY 374 LKNATGQFLNFDNTNNVNGTE-ANATEYTTGADGIITITGLKEGTYLVEKKA 432
Db 583 AETA-----YEWITDKARAITYTSNDQGFVETGLADGTYNLEETLAPAG 630
QY 433 LDNSOKVILGDCATDTTNSDLLVNPT-----VENNKGTELPSTGGIGTTT 486
Db 631 LAGNIKFFVNOQSYITGGNIDYVANSNQKDATREVENKVT-IPQTGGIG 689
QY 487 VIGAGIVLVARR 499
Db 690 MLGA-VVIMKRQ 701

RESULT 6
ID 068212 PRELIMINARY; PRT; 535 AA.
AC 068212;
DT 01-AUG-1998 (TREMELrel. 07, Created)
DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Fimbrial structural subunit.
GN FIMA.
OS Actinomycetes naeslundii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Actinomycineae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T14V;
RX MEDLINE=98187920; PubMed=9529071;
RA Yeung M.K., Donkersloot J.A., Cisar J.O., Ragsdale P.A.;
RT "Identification of a gene involved in assembly of Actinomycetes
RT naeslundii T14V type 2 fimbriae.";
RL Infect. Immun. 66:1482-1491(1998).
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY
CC AN AMIDE BOND (BY SIMILARITY).
DR EMBL; AF019629; AAC13545.1; -
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMS; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Peptidoglycan-anchor.
SQ SEQUENCE 535 AA; 56038 MW; B115867F500C1356 CRC64;

Query Match 10.3%; Score 265.5; DB 2; Length 535;
Best Local Similarity 26.8%; Pred. No. 0.00012;
Matches 151; Conservative 58; Mismatches 228; Indels 127; Gaps 29;

QY 11 VASLAP-GMAVSPVTPPIAFAET---GTTVQDTQKGYKAYKVFDAIDNANVSDSNK 66
Db 18 VLTALFLGLA-----PSAVATETPNYGNIKTDATGSLAIHKHLNGGGKDIGNPTGP 72
QY 67 DGASYLIPQKAEYKASTDFN---SLFTTTTNGRTYVTKDTSANAEIATWAKSISA 122
```

Db 73 DSKGTPV-QGVVFTAYPIITDNLKDPAGWDTISLDSKAGV--PDSACTNPAAPTGLCAHKE 129  
Qy 123 NTPPVSTVTSNDDGTEVINVSQGYVYSSTVNGAVI-----MWTSTVTPNATHIERN 176  
Db 130 GTPKVSFAT--NDEGLATITEMPQVQYALVCEITTPGNIQKAKPFVTVTPHNTA---AG 184  
Qy 177 TDATW-----GDGGKTVDDKT---YSGDTPVKYTLITYKNAVNHGTEKYVQYV 222  
Db 185 ADGOWIYDVHVPKFAISVEKSIQKGLNGYGVSLIFPVS--STAPTLOAKSYFYK 242  
Qy 223 -IKOTMPSASVVDLNBGSYEVTITDGSNITTLTGSEKATGKYNLLENNNFTITIPWA 281  
Db 243 QLRDT-----LDDRLTAVTATEVSLGCTTLDPTDYKV-----DTKGQTVTVTFT 286  
Qy 282 ATNPTFGNTQNGANDFFYKGINIITVTVGLKSGAK-----PGSADLPENTN 330  
Db 287 ABLKKIKAAAPGKVSFAVQ--KVTARNGAITNRAQVISDVTVAEQPPTPEEPANPE 344  
Qy 331 IATINPNTSNDPQKVTVRDGOITIKKID-----GSTKASLOCAIFVLKKNATQCFLNFD 386  
Db 345 ----NPTTSNE-----VTSRWGDLKKVDNHOQGDQKAGLOQAFQLYKAKNAYAG-TC 394  
Qy 387 TNNVWGTGA--NATEYTTGADGIIITIGL-----KEGT---YYLVEKKAP 427  
Db 395 TKDKEGDPAINGETTLTTDAQAINVKGLFISDSIDGANRDNQKDATARCIVLVETKAP 454  
Qy 428 LGNLLDNDKQVILGDA-----TDTNSDNLVNPVTNNKGT--ELPSTGGIGTTI 478  
Db 455 AGTVL-----PAGDGAFTPVKIEGAVTTDNV-----TIENKQSVPGPLTGTANGMLI 503  
Qy 479 FYIIGA-ILVIGAGIVLVARRLR 501  
Db 504 LTASGASLLMIAVGSVLVARYRER 527

RESULT 7  
Q97SC2  
ID Q97SC2 PRELIMINARY; PRT; 665 AA.  
AC Q97SC2;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
GN SP0463.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OX Streptococcus.  
NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC BAA-334 / TIGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Lofthus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";  
RL Science 293:498-506(2001).  
DR EMBL; AB007358; AAK74623.1; --  
DR PIR; F95053; F95053.  
DR TIGR; SP0463; --  
DR InterPro; IPR008454; Cna.B.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF05738; Cna.B; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR TIGRFAMs; TIGR01167; LPXPG\_anchor; 1.  
KW Complete proteome.  
SQ SEQUENCE 665 AA; 71364 MW; 6F2BD7616398CBA5 CRC64;

Query Match 10.3%; Score 265.5; DB 16; Length 665;  
Best Local Similarity 24.4%; Pred. No. 0.00016;  
Matches 164; Conservative 63; Mismatches 227; Indels 219; Gaps 33;  
Qy 9 LLVASLAFGMVSPVPIAFAAETGTTVODTQKGATYKAYKVFDAEIDNANVSDSNKOG 68  
Db 15 LITASLFAAAT-----VFAAGTTTSTVTHKLLATDGMKDKIANELETGNTA-GNKVG 67  
Qy 69 ASYLPQGGKEABYKASTDFNSLFTTTN-----GGRYVTVTKDTSANBIATWAKISAN 123  
Db 68 ---VLP-----ANAKEIAGVMFVWNTNNEIIDENGQTLGVNIDP-----QIFKLSGAMP 114  
Qy 124 TTPVSTVTSNDDGTEVIN--SQGYVYSV--TVNGAVIMVTSVTP----- 168  
Db 115 ATAMKLLTAEGAKFNTANLPAKYIYIHSLSYTVVGEDGATLTGSKAVPIEIELPLND 174  
Qy 169 --NATHEKNVTA---TWGDGGGKT-----VDQKT---YSGDTPVKYTLITYK--NAVNY 212  
Db 175 VVDAHVPKTEAKPKIDKDFGKANPDPTRVDKPTPNVHQGVGVVEYIVTKIPALANY 234  
Qy 213 -----HGTEKYV-----QYVIKDT-----MPSASVVDLNEG-- 238  
Db 235 ATANSDRMTEGLAFNKGTVKYTVDDVALEAGDYALTEVATGFDLKLTDAGLAKVNDQNA 294  
Qy 239 -----SYEVTTDGS-----GNITTLTGSEKATG---KYNLLENNNFTITIPWA-A 282  
Db 295 EKTVKITYSATLNDKAIIVEPESNDVTENYNNPDHNTPKPNKPNENGDLTLTKTWVDA 354  
Qy 283 TWTPTG-----NTQNGANDFFYKGINI-----TVTYTGVLSK-----G 317  
Db 355 TGAPIPAGAEATFDLVNAQTG-----KVQTVTLTDDKNTVTVNGLDKNTYEKFEVER 407  
Qy 318 AKPGSADLPENTNIATI-----NPNTSNDPQKVTVRDGOITIKKIDGSTKASIQGAI 371  
Db 408 IKGSADYQEIITTAGIYAVNWKDENPKPLDTEPKVTVYGGKFKVKNKDNK--LAGAE 465  
Qy 372 FVLKNA--TGQFL----- 382  
Db 466 FVIANADNAGQYLARKADKVSQEEKVLVTTKDALDRAVAAYNALTAQOQTQOEKEKVDK 525  
Qy 383 ---NEN-----DTNNVWGTG---ANATEYTTGADGIIITGLKSGTYLVLEKKAELGVN 431  
Db 526 AQAYNAVAIAANAFEFWADKNENVKLVSDAQGRFEITGLLAGTYTLEETKQPAGYA 585  
Qy 432 LLDNSQKV-----ILGDGATDTTNSDNLVNPVTNNKGTETLPSTGGIGTTFIYIG 483  
Db 586 LLTSRQKEVTSATSYSGGIEYTAGSK--DDATKVNKKITIPQTGGIGTTFIYAVAG 643  
Qy 484 AILVIGAGIVLVA 496  
Db 644 AAIM---GIAVYA 653

RESULT 8  
Q9X4D9  
ID Q9X4D9 PRELIMINARY; PRT; 533 AA.  
AC Q9X4D9;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
GN Type-1 fimbrial major subunit precursor.  
OS Actinomyces naeslundii.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Actinomycetaceae; Actinomycetaceae; Actinomycetes.  
OX NCBI\_TaxID=1655;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LY7;  
RX MEDLINE=99242779; PubMed=10225854;  
RA Li T., Johansson I., Hay D.I., Stromberg N.;  
RT "Strains of Actinomyces naeslundii and Actinomyces viscosus exhibit

RT structurally variant fimbrial subunit proteins and bind to different  
RT peptide motifs in salivary proteases.",  
RL Infect. Immun. 67:2053-2059(1999).  
DR EMBL: AF107020; AAD28829.1; --  
DR InterPro: IPR008454; Cna\_B.  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR Pfam: PF00738; Cna\_B; 2.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR PROSITE: PS00847; GRAM\_POS\_ANCHORING; 1.  
KW Signal.  
FT SIGNAL. 1 30 POTENTIAL.  
FT CHAIN 31 533 TYPE-1 FIMBRIAL MAJOR SUBUNIT.  
SQ SEQUENCE 533 AA; 57062 MW; F247CCACED4E3F38 CRC64;  
  
Query Match 10.1%; Score 261; DB 2; Length 533;  
Best Local Similarity 25.1%; Pred. No. 0.00019;  
Matches 146; Conservative 60; Mismatches 211; Indels 164; Gaps 29;  
  
QY 13 SLAFGMVSPVTPPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNANVSDSN--KDGAS 70  
Db 17 TLAAGALVSP--TGAAPADPNGSTI-DPQAATLTIVHK-----CEQTDITNGVKEGAG 66  
QY 71 YLIPQGEAEYKASTDFNSLFT-----TTTNGGRT-----YVTKDITASANEIATWAK 118  
Db 67 NEDPQ---AECKPVSDFEFTIKLNVLDITTDGKTLADPKGYVVRKAGALK-----STTVQ 119  
QY 119 SISANTTPVSTVTSNNDGTEVINVSQYGYVVSSTVNNGAVIM-----V 163  
Db 120 KITTGANGLASFTDAQ---TEV-----GAYLSETRTPDKVIPAEDEFVVTLPMTNPQDT 170  
QY 164 TSVTPNATIEKNTDATWGGGKTV--DQKTVSGTVKVTIITYK--NAVNYHGTETKVQY 221  
Db 171 TKWYNNVHVYPKNTLS-----GVDKQVTDKPAKSGRDIITYTITTSIPKVDYPGGARIKRY 226  
QY 222 VIKDTM-----PSASVVDLNEGSYEVTITDGSNITTLTQSEKATGKYNLLE-- 270  
Db 227 EVVDRLDKRIKKEALTPVVKIV---GQNEVLADITDITLTAEKGDHNWATIQLTEEG 282  
QY 271 -----NNN-----FTTIPWAATNPPTGNTONGANDDFYKGINITVITYGVLK 315  
Db 283 RRKASEARYNGETKLOVTLTAKFDAAVNLEGDLNTA----- 321  
QY 316 SGAKPGSADLPENTNATINPNTSNDPG---QKVTVRDQGIITIKIDG---STKASLOG 369  
Db 322 -GLIPN--DSP-NFTWDPNNPGTTDIPGIPPTPVLISKYKGVVLTGTGDLADKTKYNG 377  
QY 370 AIF-----VLKATGQFLNEND--TNNVWGTETANATEYTTGADGIITITGLK----- 415  
Db 378 AQFQVECTKTAGAMLRSDSPSTQVDPITIGKEKFTTTAGOGTVEINVLANDYVNGA 437  
QY 416 -----EGTYLVLEKAPILGYN-----LLDMSQKVLGDGATDTTNSDNLVNP 458  
Db 438 KKQDLTDEYVCLVETKAPESYSLQADPLPFRVLAKEAK-----KAATEVTITD----- 487  
QY 459 TVENNKGTETLPSTGGITTFYIIGALLVIGAGIVLVARER 499  
Db 488 -IPKNAGFRPLTGANGVIFLTAGALLVAGGAVVAYANKR 527  
  
RESULT 9  
Q9X4D8 PRELIMINARY; PRT; 533 AA.  
AC Q9X4D8;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Type-1 fimbrial major subunit precursor.  
GN FIMP.  
OS Actinomyces naeslundii.  
OC Bacteria; Actinobacteridae; Actinobacteriales;  
OC Actinomycetaceae; Actinomycetaceae; Actinomycetes.  
OX NCBI\_TaxID=1655;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=P-1-K;  
RX MEDLINE=9242779; PubMed=10225854;  
RA Li T., Johansson I., Hay D.I., Stromberg N.;  
RT "Strains of Actinomyces naeslundii and Actinomyces viscosus exhibit  
structurally variant fimbrial subunit proteins and bind to different  
RT peptide motifs in salivary proteases.",  
RL Infect. Immun. 67:2053-2059(1999).  
DR EMBL: AF107019; AAD28828.1; --  
DR InterPro: IPR008454; Cna\_B.  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR Pfam: PF00738; Cna\_B; 2.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR PROSITE: PS00847; GRAM\_POS\_ANCHORING; 1.  
KW Signal.  
FT SIGNAL. 1 30 POTENTIAL.  
FT CHAIN 31 533 TYPE-1 FIMBRIAL MAJOR SUBUNIT.  
SQ SEQUENCE 533 AA; 56772 MW; 6261AC963AD1D75C CRC64;  
  
Query Match 10.0%; Score 259; DB 2; Length 533;  
Best Local Similarity 26.1%; Pred. No. 0.00024;  
Matches 148; Conservative 67; Mismatches 216; Indels 136; Gaps 31;  
  
QY 13 SLAFGMVSPVTPPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNANVSDSNKDGASYL 72  
Db 17 TLAAGALVAP--TGAAPADPNGSTI-DPDAATLTIVHKCEQTDITNGVKEGTGNE----- 69  
QY 73 IPQGEAEYKASTDFNSLFT-----TTTNGGRTV--TKKDTASANEIATWAKSISANTT 125  
Db 70 -PQ---AECKPVSDFEFTIKLNVLDITTDGKTLADLKGDVVKAG--ALKSDTVQKIIT 123  
QY 126 PVSVTESNNDGTEVINVSQYGYVVSSTVNNGAVI-----MVT--SVTP-----NA 170  
Db 124 GANGLASFTDAQTEV-----GAYLSETRTPDKVIPAEDEFVVTLPMTNPQDTAKWYNV 177  
QY 171 TTHEKNTDATWGGGKTV--DQKTVSGTVKVTIITYK--NAVNYHGTETKVQYVVIKDTM-- 227  
Db 178 HVPKNTLS-----GVDKQVTDKPAKSGRDIITYTITTSIPKVDYPGGARIKRYEVVDRLD 233  
QY 228 -----PSASVVDLNEGSYEVTITDGSNITTLTQSEKATGKYNLLENNFTITI 278  
Db 234 KRKKEALTPVVKIV---GQNEVLANTCT-DYTLIT-----AEK-----DHNWATIQ 278  
QY 279 PWAATNPPTGNTONGANDDFYKGINITVITYGVLS-----GAKPGSAD-LPENTNIA 332  
Db 279 TEEGRKASEARYNGND-----TKIEVTLTAKLDAAVNLEGDLSTAGFIPNDSPNF 331  
QY 333 TINPN--TSNDPG--QKVTVRDQGIITIKI---DGSTKASLOCAIFVL---KNATG 379  
Db 332 TMDPNPGTTDIPGIPPTPVLISKYKGVVLTGTGDDLDKTKYNGAQFQVECTKTASG 391  
QY 380 QFLNFD--TNNVWGTETANATEYTTGADGIITITGLK-----EGTYLV 422  
Db 392 ATLSDSPSTQVDPITIGKEKFTTTAGOGTVEINVLANDYVNGAKKQDLTDEYVCLV 451  
QY 423 EKAPILGYN-----LLDMSQKVLGDGATDTTNSDNLVNPVNNKGTETLPSTG 472  
Db 452 ETKAPEYGNLQADPLPFRVLAKEAK-----KAATEVTITD-----VPKNAGFRPLTG 500  
QY 473 GIGTTFYIIGALLVIGAGIVLVARER 499  
Db 501 ANGVIFLTAGALLVAGGAVVAYANKR 527  
  
RESULT 10  
Q9CHH3 PRELIMINARY; PRT; 614 AA.  
AC Q9CHH3;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Unknown protein.  
GN YHGE OR LL0758.





```

ID Q48707 PRELIMINARY; PRT; 507 AA.
AC Q48707;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Orf2 protein.
DE Orf2 protein.
GN ORF2.
OS Lactobacillus leichmannii.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=28039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 20076;
RA Schenk-Groeninger R.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X81869; CAA57459.1; -.
DR PIR; S52348; S52348.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
SQ SEQUENCE 507 AA; 53931 MW; 2DF1C7B56C720CDF CRC64;

Query Match 9.7%; Score 251; DB 2; Length 507;
Best Local Similarity 24.7%; Pred. No. 0.00052;
Matches 139; Conservative 76; Mismatches 204; Indels 142; Gaps 27;

QY 1 MKKK-----MIQSLIVASLAFGMVSPVPIFAAETGT---ITV----- 37
Db 1 MKSKLSLITGLMILPLLSPATPKVS-----AADTNSVKVTLHKRVFDSAQEA 53

QY 38 -QDTOK-----GATYKAYKVFDAEIDNANVSDSNKDGASLYLIPQGEAKYKAST 85
Db 54 KQMTGEIMNDFGTGLNGVTFKAYNVTDHLSLRKSDSAQDAVTAIQSDAKSDNLPSY 113

QY 86 DFNSLFTTTNGGRTYVTKK--DTASANEIATWAKSISANTTPVSTVTSNNDGTEVIN- 142
Db 114 AGSAIATETT-----ATSKGEDGIAAFDNLNKLKSDGNYQTYLFLVETDSPTDVTQQAAP 167

QY 143 -VSQGYGVYSSVYNGAVIMVTSVFNATIEHKNTDATWGDGGGKTVDQK-----TY 194
Db 168 IVLTPYIKTSDT-----SAINHDIOIYPQNVKST---PITKDLDEASKKDLAVTL 215

QY 195 SVGTGVKTYITYKNVNVHGTGVK-YQYVIKDMPSASVVDLNEGSYEVTITDGSNGIIT 253
Db 216 PDGSTI-YNAQYKSGFYNTVNPWNIKDKTF--NVVDKPTGIDI---DAS---TV 265

QY 254 LTQSEKATGKYNLLENNTTITIPWAAINTPTGNTONGANDDFFYKGINITVTVTG 313
Db 266 SIDGLTKST-DYTVNKKDNGYQVVF-----KTTSAAVQALAG-----KSLTITYKAT 311

QY 314 LKSGAKFGSA-----DLPENTNIATINPNTSDPDQKVTVRDGOITIKKIDGSTKAS 366
Db 312 LTNNATPDKAIGNTATISGNNTNI-----TSPANGPRIYTGGAQF-VAK-DSQNK 363

QY 367 LQGAIFVLKNAFGOLFNFNDNNVEWTEAN-----ATEYTTGADGIITITGL 414
Db 364 LAGAEFQLVKVD-----SNGNIVSYATQASDGSYTWNSDASATEATTYTSANGVALKGL 417

QY 415 KEG-----TYYLVEKKAPLGYNLLNSQVILGDGATDTTNSDNLNVPNTVENNKGE 467
Db 418 SYSKLDGSGESYALLEIQAPDGYAKLDSPVKFSITQSGFSDN-----KITIDNTKEGL 471

QY 468 LPSTGGIGTTTIFYIIGAILVIGA 490
Db 472 LPSTGGKGIYIFLAIGVIMIVA 494

RESULT 13
ID Q9X4D2 PRELIMINARY; PRT; 533 AA.
AC Q9X4D2;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Type-1 fimbrial major subunit precursor.
DE FIMP.
GN FIMP.
OS Actinomyces naeslundii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B-1-K;
RX MEDLINE=99242779; PubMed=10225854;
RA Li T., Johansson I., Hay D.I., Stromberg N.;
RT "Strains of Actinomyces naeslundii and Actinomyces viscosus exhibit structurally variant fimbrial subunit proteins and bind to different peptide motifs in salivary proteins.";
RT Infect. Immun. 67:2053-2059(1999).
DR EMBL; AF106035; AAD28827.1; -.
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR SIGNAL.
KW SIGNAL.
FT SIGNAL 31 533 TYPE-1 FIMBRIAL MAJOR SUBUNIT.
FT CHAIN 1 533
SQ SEQUENCE 533 AA; 56816 MW; D65207B99B24528E CRC64;

Query Match 9.7%; Score 250; DB 2; Length 533;
Best Local Similarity 24.7%; Pred. No. 0.00061;
Matches 142; Conservative 65; Mismatches 216; Indels 152; Gaps 30;

QY 13 SLAFGMVSPVPIFAAETGTITVQDTQKATYKAYKVFDAEIDNANVSDSNKDGASYL 72
Db 17 TLAAGALVAP-TGAAPADPNGSTI-DPDAATLTVHKCEQDTDNGVKEGTGNE- 69

QY 73 IPQGEAKYKASTDFNSLFT-----TTNGGRTVY-TKQDTASANEI-ATWAKSISANT 124
Db 70 -PQ--AECKPSVDEFTITKLNVDLTYDGMKTLADLKGVDVVKAGALKSTTVQKITGA 125

QY 125 TPVSTVTSNNDGTEVINVSQYGVYSSVYNGAVI-----MVT--SVTP-----N 169
Db 126 NGLASTDAQ---TEV-----GALVSETRPDKVIPAEDFVTLPMNPDQTAKNYN 176

QY 170 ATIEHKNTDATWGDGGKTV-DQKTVSVGDVTKYITYK-NANVHGTGVKTVQYVTKDM 227
Db 177 VHVYPKNTLS---GVDKQVTDKPAKPSGRDITYTITTSIPKVDYPPGARIKRYEVDRL 232

QY 228 -----PSASVVDLNEGSYEVTITDGSNGIITLTQSEKATGKYNLLE- 270
Db 233 DKRIKKEALTPVVKIV---GQNEVTLADTDTYTLITAEKHNWATIQLTEGRKASE 288

QY 271 ---NN-----PTITPWAATNPTGNTONGANDDFFYKGINITVTVTVGLKSGAKPG 321
Db 289 ARYNGETKQLVTLTAKFDAANLEGLDLSNTA-----GLIPN 326

QY 322 SADLPENTNIATINPNTSDPDG---QKTVTRDGOITIKKIDG---STKASLOGALFVL- 374
Db 327 --DSP-NFTWDPNPNPGTTTIDPGITPTPVLSTKYKVLTKTGTQDQLADTKYNGAQFOVY 383

QY 375 ---KNATQGLNFD--TNNVEWGTGANATEVTTGADGIITITGLK----- 415
Db 384 ECTKTASGATVRSDPSPTQVDPVTIGGKFTTACGGVTEINLYRANDYVNGAKDKDLT 443

QY 416 -ECTYVLVEKKAPLGYN-----LLDNSQKVLGDGATDTTNSDNLNVPNTVENNK 464
Db 444 DEDYICLVETKAPEGYSLOADPLFPVLAEKAEK---KAAETEVTVD-----IPQNA 492

QY 465 GTELPSTGGIGTTTIFYIIGAILVIGAVLVARRR 499
Db 493 GFRPLPLTGANGVIFLTVAGALLVAGAVVAYANKR 527
```

Db 430 TTGQLTVSGLTAGNYQFVETKAPSGYQLNNAAPLSFTIKPNQTAVVVTAATDEPVTPEGTT 488

QY 457 NPTVENNKGTPLPS--TGGIGTT 477

Db 490 EPSKPGEPCTTEPSKPGEPGTT 511

RESULT 15

Q8B6E4 PRELIMINARY; PRT; 554 AA.

AC Q8B6E4; Q8B6E4;

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Hypothetical protein.

GN G8S0628.

OS Streptococcus agalactiae (serotype III).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=216495;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MEM316 / Serotype III;

RX MEDLINE=2242508; PubMed=12354221;

RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L., Meadek T., Zouine M., Couve E., Lalloui L., Poyart C., Trieu-Cuot P., Kunst F.;

RA "Genome sequence of Streptococcus agalactiae, a pathogen causing

RT invasive neonatal disease.";

RL Mol. Microbiol. 45:1499-1513 (2002).

DR EMBL; AL768846; CAD46272.1; -.

DR Sagal18; gb80628; -.

DR InterPro; IPR001899; Gram\_pos\_anchor.

DR InterPro; IPR008454; Cna\_B.

DR Pfam; PF05738; Cna\_B; 1.

DR Pfam; PF00746; Gram\_pos\_anchor; 1.

DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 554 AA; 60168 MW; 644984B49729AD1 CRC64;

Query Match 9.5%; Score 244.5; DB 16; Length 554;

Best Local Similarity 25.4%; Pred. No. 0.0011;

Matches 151; Conservative 61; Mismatches 218; Indels 165; Gaps 31;

QY 1 MKKMIQSLVASFMAVSPVTFIAFAETGTTV-----QDTQKATYKAYK--- 51

Db 3 LSKLLFLFAAULTVAGSTVEPVAQFA---TGMISIVRAAEVSQERPAKTTVNYKLD 58

QY 52 -FDAIDNANVSDSKOG--ASYLIPQCKEAEYKASTDFNSLFTTTTNGGRTYVTKD 108

Db 59 SYKSEI--TSNGIENKDGIVSNVAKGDNVKGQGVQF-----KRYKVD-I 105

QY 109 SANEIATWAKSISANTTPVSTVTE-----SNDGTEV---INVSQGYVYVSSVTNN 157

Db 106 SVDELKK-LTTVEADAKVILEBGVSLPKQTKAQLVLDALOSKNVRLYVEDLKN 164

QY 158 GAVIMVTSVTP-----NAT-----IHKR--TDATWGGGGKTVQDTYSGVD 198

Db 165 PSNITKAVAPVPELPELVANSTGTGFLSEINIPKVVVTDPEKTKDVKLQ-----D 218

QY 199 TVKYITITYKNAVNYHGTEKVVQYVVKDTPMSASVVDLNEGSVE-VTITDGSNGITTLTQ 257

Db 219 DAGYII-----GEE--FWFLKSTIPA-----NLGYEKFEITDKPADGLTY--- 258

QY 258 SEKATGKYNL-----LENNNETITIPWAAATNPTGNTQNGANDDF-----FYKGI 303

Db 259 --KSVGKIKIGSKTLNRDEHYTI-----DEPTVDNQNTLKITFKPEKFEIAELLKGM 309

QY 304 NITITYTVGLKSGAKPGSADLPENTNIATI-----NPVASTINEKAVLKAIENTFELQYDHTPDKADNP 337

Db 310 -TLVNQDALKATANTDAAFLPEIPVASTINEKAVLKAIENTFELQYDHTPDKADNP 368

```

QY 338 TSNDPQGVTRDGOITIKKIDGSKASLOGAIFVLKNATQGFNLNNDNNVWEGTEAN 397
Db 369 PSN--PPRPEVHTGGRFVKKDSFTQTLLGAEFDLLASDGTAVKWTDA-LIKANTNKN 425
QY 398 --ATEYTTG-----ADGIITITGL-----KEG-----TYLVEKKAFLGYNLLDNSQ 437
Db 426 YIAGEAVTGOPIKLKSHGDTGTFEIKGLAYAVDANAEGTAVTYKLKETAPEGYVIPDKBI 485
QY 438 KVLGDGA-----TDTTNSDNLVNPTVNNKGTLPSTGGIGTTIFYLIGAILV 487
Db 486 EFTVTSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTGGIGTAIFVAIGAAMV 540

RESULT 16
Q820S9 PRELIMINARY; PRT; 554 AA.
AC Q8E0S9;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cell wall surface anchor family protein.
GN SAG0845.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae."
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014220; AM99537.1; -.
DR TIGR; SAG0645; -.
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
KW Complete proteome.
SQ SEQUENCE 554 AA; 60168 MW; 64498D4B49729AD1 CRC64;

Query Match 9.5%; Score 244.5; DB 16; Length 554;
Best Local Similarity 25.4%; Pred. No. 0.0011;
Matches 151; Conservative 61; Mismatches 218; Indels 165; Gaps 31;

QY 1 MKKKMIOSLLVASLAFGNVSPVTPPIAFAETGTTV-----QDTQKGATYKAYKV--- 51
Db 3 LSKKLLFSAAVLTVAGSTVEPVAQFA---TGMISIVRAAEVSQERPAKTVNIYKLOAD 58
QY 52 -FDAIDNANVSDNSKDG--ASYLIPQKGEAYKASTDPNSLFTTTTNGGRTYVTKKDTA 108
Db 59 SYKSEI-TSNGGIENKQGEVSNYAKLGDNVKGGQGVQF-----KRYKVKTD-I 105
QY 109 SANEIATWAKSIGISANTTTPVSTVTE-----SNNDGTEV---INVSYQYGVYVSSVTNN 157
Db 106 SVDELKK-LTTVEADAKVGTILEEGVSLPKQTNAGVLVDALDSKNVRVLYVEDLKNs 164
QY 158 GAVIMVTSVTP-----NAT-----IHEKN-----TDATWGDGGGKTVQDKTSVGD 198
Db 165 PSNITKAYAVFPVLELPVANSTGTGFLSEINIYPRNVVTDPEKTDKQVKLQG-----D 218
QY 199 TVKYTIITKNAVNYHGTKEVQYVYIKDTPMPSASVVVDLNEGSYE-VTITDGSNITLTITQ 257

```

```

Db 219 DAGYTI-----GEE--FKWFLKSTIPA-----NLGDYKFEITDKPADGLTY--- 258
QY 258 SEKATGKYNL-----LENNNFTIIPWAAATNPTGTQNGANDDF-----FYKGI 303
Db 259 --KSGKIKIGSKTLNRDEHYTI-----DEPTVDNQNTLKITFKPKPKFAIELLKG 309
QY 304 NTITVTVTVGLKSGAKPGSADLPENTNIATI-----NPN 337
Db 310 -TLVKNQDALDKATANTDDAAFLPVPVASTINERKAVLGKAIENTFELQYDHTDPKADNPK 368
QY 338 TSNDPQGVTRDGOITIKKIDGSKASLOGAIFVLKNATQGFNLNNDNNVWEGTEAN 397
Db 369 PSN--PPRPEVHTGGRFVKKDSFTQTLLGAEFDLLASDGTAVKWTDA-LIKANTNKN 425
QY 398 --ATEYTTG-----ADGIITITGL-----KEG-----TYLVEKKAFLGYNLLDNSQ 437
Db 426 YIAGEAVTGOPIKLKSHGDTGTFEIKGLAYAVDANAEGTAVTYKLKETAPEGYVIPDKBI 485
QY 438 KVLGDGA-----TDTTNSDNLVNPTVNNKGTLPSTGGIGTTIFYLIGAILV 487
Db 486 EFTVTSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTGGIGTAIFVAIGAAMV 540

RESULT 17
Q9X4D1 PRELIMINARY; PRT; 535 AA.
AC Q9X4D1;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Type-1 fimbrial major subunit precursor.
GN FIMP.
OS Actinomyces viscosus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1656;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19246;
RX MEDLINE=99242779; PubMed=10225854;
RA Li T., Johanson I., Hay D.I., Stromberg N.;
RA "Strains of Actinomyces naeslundii and Actinomyces viscosus exhibit
RA structurally variant fimbrial subunit proteins and bind to different
RA peptide motifs in salivary proteins."
RL Infect. Immun. 67:2053-2059(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19246;
RX MEDLINE=99242779; PubMed=10225854;
RA Li T.;
RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF106034; AAD28826.2; -.
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 535 TYPE-1 FIMBRIAL MAJOR SUBUNIT.
SQ SEQUENCE 535 AA; 56378 MW; E64F5B13A00F1E7 CRC64;

Query Match 9.4%; Score 243; DB 2; Length 535;
Best Local Similarity 23.6%; Pred. No. 0.0013;
Matches 133; Conservative 71; Mismatches 232; Indels 128; Gaps 25;

QY 13 SLAFGNVSPVTPPIAFAETGTTITVQDTQKGATYKAYKVFDAIDNANVSDNSKDGASYL 72
Db 17 TLAVGALIAP-TGAAAPADPNGSTIDPETH-ATLTVHKCEQTDNNGVKEGTGNEDEPTT-- 72
QY 73 IPOKGEAYKASTDFN-----SLFTTTTNGGRTYVTKKDTASANEIAT-WAKSISANTTP 126
Db 73 -----ECNPVSGVEFTITRLNVDLTDEGKTLAALQGNVAGQAUKTQTSITTYTGADG 127

```

```
QY 127 VSTVTESNDGTEVINVSQYGYVYSSVTVNGAVIMVTS--VTPNATHEKNTDATWG-- 182
DB 128 LASFTDAQTD-----VGAYLVSETPKVIPAEDEFVVTLPMTNPEKTEWYNVH 178
QY 183 -----DGGGKTVDQKTVS--VGDVVKYITIK--NAVNVHGTGKVVQYVVKDTM----- 227
DB 179 VPKTVSVGDVQVSDKQVSGSGNDITYTITTSIPKVDYAGGARIKRYEVVDOLDKRIKK 238
QY 228 -----PSASVVDLNEGSYEVTITDGGSNITTLTQSGSEKATGKYNLLSENNFTITIPWAAT 283
DB 239 DQLTPVVKIIGAAGNPEITLVDGTDYTVITADGANHNWATIQLTTEGRKAA-----EAR 294
QY 284 NTPGTONTGANDDPFKGINITYTIT-----GVLSGAKPGSGADLPENTNIAT 333
DB 295 ANGTGETK-----VQVTLTAKFSDVDLEGLTSLNTAGLIPSDSP--NPDMDP 339
QY 334 INPNTSNDPPGQKVT---VRDGOITIKKI---DGS TKASLQGAIFVL-----KNATGOQLN 383
DB 340 SNPGT--DVPGTPPTTSISKYKVNLTGTGNLADTKYNGAEFQVYECTKTATGATLK 397
QY 384 FND---TNNVWGTEANATEYTTGADGIIITIGLK-----EGTYILVEKKA 426
DB 398 DADASTADKVDPLTIGGKQTFTEGGQMVSIATLRANDVYNGAEKQLTDDYYCLVETKA 457
QY 427 PLGYNL-----LDSKVLGDGATDTTNSDNLNVPNTVNNKGTBLPSTGGIG 475
DB 458 PEGYTLQADPPIPRVLAADAEEKI-----PTEVTVTD-----VPKAGFRLPLTGANG 505
QY 476 TTFIVFIIGAILVIGAGIVLVARRR 499
DB 506 VIPLTIAGALLVAGGAVAYANKR 529

RESULT 18
QBE4C4 PRELIMINARY; PRT; 674 AA.
AC QBE4C4;
DC QBE4C4;
DT 01-MAR-2003 (TremBrel. 23, Created)
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
DE Hypothetical protein.
GN GBS1477.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunat F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease."
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766851; CAD47136.1; --
DR Sagaliet; gbs1477; --
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00738; Cna_B; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFam; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; P550847; GRAM_POS_ANCHORING; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 674 AA; 73218 MW; 211133B59FC4727D CRC64;

Query Match 9.0%; Score 232.5; DB 16; Length 674;
Best Local Similarity 21.4%; Pred. No. 0.0049;
Matches 149; Conservative 82; Mismatches 217; Indels 249; Gaps 32;
QY 9 LLVASLAFGMVSPVTPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNANVSNDKDG 68
```

```
DB 16 LLTSL---FSVAP-----AFADDVTTDTV-----TLHKIWMPOAAFNFTEGTGKND 61
QY 69 ASYLIPOGKE-----AEYKASTDFNSLFTTTTNGGRYVTKKDTASANEIATWAKSISANT 124
DB 62 SDYV---GQINDLSYFGSDAKEI-----KGAFFVFKNETGT-----KFTENG 104
QY 125 TPVSTVTSNSNDGTEVIN-----VSQYGYVYV-----SSTVNGAVIMVTSV 166
DB 105 KEVDTLEAKDABGGAVALSGLTKDTGPAFNATKLKGTQIYVELKEKSNYDNNGSILADSKA 164
QY 167 TP-----NATHEKNT-----DATWGGG-----GKTV-DQ 191
DB 165 VEVKITLPLVNNQGVVVKOAHYIPKNTETPKQVDKFNADKOLDYDNRKDKGVVSATVGGK 224
QY 192 KTVSYG-----DTVKYTYIKNAV-----NYH--GTEKYVQYV 222
DB 225 KEYIVGTLKGSYDKLVWTDSTWTKGLTFNNNVKVLTDGKDFPVLYNKLVTDDQGFRLA 284
QY 223 IKDTPSASVVDLNEGSYEVTIT-----DGSNIITLTQSGSEKATGKYNLLSE----- 270
DB 285 LNATGLAAVAAAADKDVIEIKITYSATVNGSTTVEVPETNDVKLDYGNPTSESEPEQGT 344
QY 271 --NNFTITIPAAATNPT-----GNTQNGAND-----DFFYKG 302
DB 345 PANQEIWKIDWADGTITDNNVAVKAITLQEKQTDGTWVNVASHEATKPSRFEHTFTG 404
QY 303 INTIITVTVTLK--SGAKPGSADLPENTNIATINPTNSNDP-----CQKVT 348
DB 405 LDN--TKTYRVVERVSGYTPYVSP--KNGVTVIKNNKNSNDPTPINPSPKVVTYGRKP-- 460
QY 349 VRDGOITIKKIDGST-----KASLQGAIFVLKNATGOQLN----- 383
DB 461 VKTNQANTERLAGATFLVKKEGKYLARKAGATAEAKAAVKTAKLALDEAVKAYNDLTKE 520
QY 384 -----FND-----TNNVWEGTEA---NATETTTGADGIIITIGLK 415
DB 521 KOEGEGKTALATVDQKAYNDAPVKANYSEWADKKADNVVKLIISNAGQGFETGLD 580
QY 416 EGTYYLVEKKAPLGYNLNLSOKVILG-----GATDTTNSDNLNVPNTVE-- 461
DB 581 KGTYSLEETQAPAGYATLS-----GDVNFVETATSYSGKATTDIADKSGVKDAQOV 633
QY 462 NNKGTELPSTGGIGTITFIYIIGAILVIGAGIVLVARR 498
DB 634 QNKVVTIPQGTGGIGTITFIYIIGAILVIGAGIVLVARR 498

RESULT 19
QBYAG7 PRELIMINARY; PRT; 793 AA.
AC QBYAG7;
DC QBYAG7;
DT 01-MAR-2002 (TremBrel. 20, Created)
DT 01-MAR-2002 (TremBrel. 20, Last sequence update)
DE Putative peptidoglycan bound protein (LPXTG motif).
GN LMO0159.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
ON NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dubsurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
```

```
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehlant J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL591973; CAC98374.1; -.  
DR PIR; AH1094; AH1094.  
DR ListiList; LMO00159; -.  
DR InterPro; IPR008454; Cna B.  
DR InterPro; IPR008456; Collagen_bind.  
DR InterPro; IPR001899; Gram_pos_anchor.  
DR Pfam; PF05738; Cna B; 4.  
DR Pfam; PF00746; Collagen_bind; 1.  
DR Pfam; PF00746; Gram_pos_anchor; 1.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 793 AA; 84506 MW; F3B69B074884FIAC CRC64;  
  
Query Match 9.0%; Score 232; DB 16; Length 793;  
Best Local Similarity 25.1%; Pred. No. 0.0662;  
Matches 125; Conservative 66; Mismatches 189; Indels 118; Gaps 23;  
  
QY 45 TYKAYKVF--DAIDNANVSDSNKDGASYLIPO-----GKEAYK---- 82  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|  
239 TYSADRVFTAGAPISSTNFS-ATSDGFSVALGNLTDVQISYITTTTDDGKSKTDYDNTAK 297  
QY 83 -ASTDF--NSLFTTTTNG-----GRYVTKD--TASANELATW-----A 117  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
298 LAGTDFVTKQTSWTWTFASGGGGGGTGTGVLTKEDAKTLEGAFLVDKSGTVLQ 357  
QY 118 KSIASANTPVSVTESNNNDTEVINVSQGVYVSST-----VNGAVIMVTSVTPNA 170  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
358 ENITNASGQSLADLKFDYQLIETKAPTGYKLDTPVETIGENNQAI---TVTKEN 413  
QY 171 TIHKONTATGDDGGK-TVDQKTSYSGDVTKTIYKNAVNYHTEKV-----YQV 222  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
414 TLNTGSVELTKDAATKATLAGATFELQDEKGNLTQDLKTDENGVLKVDLPVGSYQV 473  
QY 223 IKDTPMSASVVDLNEGSEYVITIDGSGNITLTQSEKATGKYNLLENNFTITIPWAA 282  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
474 -ETSAPGYKLDNPSVFEV-IGETDQVVKV-----KENTLEVG-----SVELTK 518  
QY 283 TNTPTGNTONGANDFFYKGINITVY-----TGVLK-SGAKPGSADLPEN-----TNIA 334  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
519 LDSATKATLAGATFELQDEKGNLTQDLKTDENGVLKVDLPVGSYQVETSAPTGYKLD 578  
QY 335 NPNTS-----NDPQGVTVRD-----GQITIKKIDGSTKASLOGAIFVLKNATGQFLNF 384  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
579 NSPVSEFVAGETDQVVKVTKENTLEVSGVELTKDLSATKATLAGATFELQDEKGNLTQ- 637  
QY 385 NDTNNVWEGTEANATEYTTGADGIITITGLKEGYLYLVEKKAPLGYNLNDNSOKVILGDG 444  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
638 -----TGLTIDENGVLKVDLPVGYQFVKETKAPIGVELDTPVSEFVAG 683  
QY 445 ATD---TTNSDNLVNPT 459  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
684 ETDPVVKVTKENTLVPPT 701  
  
RESULT 20  
QY591 ID QY591 PRELIMINARY; PRT; 1612 AA.  
AC QY591;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative peptidoglycan bound protein (LPXG motif).  
GN LMO2178.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-e / Serovar 1/2a;  
  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Vaquerio F., Berche P., Bloescher H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gantier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournan A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehlant J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL591982; CAD00256.1; -.  
DR PIR; AB1347; AB1347.  
DR ListiList; LMO02178; -.  
DR InterPro; IPR008454; Cna B.  
DR InterPro; IPR008456; Collagen_bind.  
DR InterPro; IPR001899; Gram_pos_anchor.  
DR Pfam; PF05738; Cna B; 6.  
DR Pfam; PF05737; Collagen_bind; 1.  
DR Pfam; PF00746; Gram_pos_anchor; 1.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1612 AA; 175072 MW; 200FBD42D03817AD CRC64;  
  
Query Match 8.5%; Score 219; DB 16; Length 1612;  
Best Local Similarity 24.2%; Pred. No. 0.054;  
Matches 116; Conservative 47; Mismatches 170; Indels 146; Gaps 23;  
  
QY 43 GATYKAYKVFDAID---NANVSDSNKDGASY--LIPOG-----KEAYK--ASTD---- 86  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
719 GGYKSYNSTTGNLDWIVSNAMAKYDNLIFDDALPTGLTYVEGSLQYRNASTSEMN 778  
QY 87 ----FNSLFTTTNGGRTVTKDTASAN---EIATWAKS---ISANTTPSVTSTVSSND 136  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
779 LVTLPLNSVGTAVTKGDKNYPTKVDTTGNKLHLEFANLENSRVFIKYTKP----- 828  
QY 137 GTEVINVSQGVYVYSSTV---NNG-----AVIMVTSVTPNATIHKEKNTDATWG 182  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
829 ----NENWYFYSYVONTAKVSDNGVGEKSYQAYASKLFNATKTAIDP----- 875  
QY 183 DGGGKTVDKQTSVGTGVTKYTIYKNAVNYHTEKVYQYVVKDTPMSASVVDLNEGSEYV 242  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
876 -----SFDKNVAVITLN---NISADRPINNPTITDTMKT-----CTTGA 912  
QY 243 TTDGSGNTTLTQSEKATGKYNLLENNFTITIPWAAATPTGNTONGANDFFYK 302  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
913 QVVKSSFKVMNETTGEDIDSKYDYITTDNNFTIQF-----DYKA 953  
QY 303 INTITVYTVGV-LKSG-----AKPGSADLPENTNIATINPNTSNDPQGVK-----T 348  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
954 TAPIKVTYITISLMSGLVSNATATASPDYGSIP-----MYKSRTTISIPAFITGSGSGT 1008  
QY 349 VRDQGITIKKIDGSTKA-SLOGAIFVLKNATGQFLNFNDNNVWEGTEANATEYTTGADG 407  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
1009 ATIGSLEITKVDKDKNSKKLTGAKFQLYTLEGD-----KAGQEA-----TTDSOG 1053  
QY 408 IITITGLKEGYLYLVEKKAPLGYNLNDNSOKVILGDGATDNTNSDNLVNPTVENNKGT 466  
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:|  
1054 KIVMDGLOGSKYKLVETEAPTGYTISDEYK-----DGKEITVTAD-IATSVTIENTEQ 1106  
  
RESULT 21  
QY5H7 ID QY5H7 PRELIMINARY; PRT; 562 AA.  
AC QY5H7;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative peptidoglycan bound protein (LPXG motif).
```



```
Db 952 -DYKATDPIKVTYSTVSLMSGLISNTATTSPDYGSLPMSYKSRRTNISPAFTIGSGGT 1010
QY 349 VRDQOITIKID--GSTKASLQGAIFVLKXNATGQFLNFDNNTNVEWGTETANATEYTTGAD 406
Db 1011 ATIGSLBITKVDKNDTK-KLTGAKFQLYTPEGD-----KAGQEA-----TDS 1054
QY 407 GIITITLKEGTYYLVEKKAPGLYNLLDNSQKVLGDGATDTTNSDNLNPTVNNKGT 466
Db 1055 GKIWMGLSQKYLVEAFTGVTISDEYK----DGKEITITAD-VATSVTIENTEQT 1108

RESULT 23
Q89PB9 PRELIMINARY; PRT; 3441 AA.
AC Q89PB9;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE BIL3563 protein.
GN BIL3563.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegasawa K., Iriuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005948; BAC4828.1;
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004601; P:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR001343; Hemlyen_Ca_bind.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00353; hemolysin_cabind; 7.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00435; PEROXIDASE_1; 2.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 3441 AA; 342613 MW; D7175RCB122D0B82 CRC64;

Query Match 8.4%; Score 215.5; DB 16; Length 3441;
Best Local Similarity 23.6%; Pred.No. 0.18;
Matches 115; Conservative 60; Mismatches 165; Indels 147; Gaps 25;

QY 27 AFAAETGTTIVDTQKATYKAYKVPDAEIDNANVSDSKDGASYLIPQKKEAYKASTD 86
Db 2200 SFNAATGTLTGSSSVANVQT--ALDS-VTYLNTSD-NPSGLARTV----- 2242
QY 87 FNSLFTTTTNGRYYTKDTSANAEIATWAKSISANTTPVS-----TVTESNND 136
Db 2243 -----TITNDG-----TANSVAV---TDTINVPVNDPVGVDNVTGSATDAGGT 2284
QY 137 GTEVINVSQYGYIYVST-VNNGAVIMVTSVTNATIHENKNTDATWGDGGGKTVOKTYS 195
Db 2285 NGTAGSNATGNVLTNDTVDNNTASLVSAIRTGAVEGSGTAGTLG----- 2331
QY 196 VGDVTKYTYTKAVNVHGTKEV-----YQYVLKDTMPSASVVDLNEGSVEYVITDGSN 250
Db 2332 -----SILVAGHTLTLANNGSYIYVND--DSAVQALNSQ---TITD-SFN 2374
QY 251 ITTLTGSEKATKYNLLENNFTITI PWAATNPTGTONGANDDFYKGINITIVTY 310
Db 2375 YAVKDPGN-----LTDTAULTVTI-----NGAND-----APNVTV--- 2404
QY 311 TGVLKSGAKPGSADLPENTNIATNP-----NTSNDPQGQ-----KVTVDGQITIKKIDG 361
```

```
Db 2405 -----PGTQEVVQNTNV-TFNGAKLISISDVBVGAGTETVTLVAHGTULTSGTTG 2454
QY 362 SYKASLQGAIFVLKXNATGQFLNFDNNTNVEWGTETANATEYTTGADGIIITIT-----GLKE 416
Db 2455 LSFTTGDGTTDATTMTESGTAANNALN---GLLYNPDTDFVGAD-TLTIITTTDQGLSD 2510
QY 417 GTYLVVEKKAPGLYNLLDNSQKVLGDGATDTTNSDNLNPTVNNKGTETLPSTGGIGT 476
Db 2511 SDTIVINQSPNPGITTTSTSDVIFYASGNTVNTATNLIN-----GTD-SITGGTGT 2562
QY 477 TIFYIIG 483
Db 2563 DTLIVTG 2569

RESULT 24
Q92FA4 PRELIMINARY; PRT; 681 AA.
AC Q92FA4;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Surface anchored protein (LPXTG motif).
GN LINO202.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkak G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordieik G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species."
RL Science 294:849-852(2001).
DR EMBL; AL596163; CAC95435.1;
DR PIR; AC1458; AC1458.
DR Listibet; LINO202;
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR008456; Collagen_bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 3.
DR Pfam; PF05737; Collagen_bind; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
KW Complete proteome.
SQ SEQUENCE 681 AA; 73391 MW; C9B713CFE1547FD0 CRC64;

Query Match 8.3%; Score 214.5; DB 16; Length 681;
Best Local Similarity 22.0%; Pred.No. 0.033;
Matches 110; Conservative 56; Mismatches 194; Indels 141; Gaps 16;

QY 17 GMAVSPVTPIAFAAETGTTIVDTQKATY-KAYKVFAE-----IDNANVSDSNK 66
Db 12 GLAFLSVLIATVQTIVKAATSYGSEFLNTVELLDKGTGPKDFGYDNNVH----- 67
QY 67 DGASYLIPQKKEAYKASTDF-----NSLFTTTNGRYYTKDTSANAEIATW 116
Db 68 --YTWISPNSTNVKAGDSMDPALPSQLATLAFNVKDSKQTVGTATVKRATNQV--- 122
QY 117 AKSISANTTPVSTVTESSND-----GTEVINVSQYGYIYVSTVN-NG 158
Db 123 -----TVVPSDYVEKHSIKGELDFWTFNQKVTGNEKINLE---FPIENSTINVDV 172
```



QY 159 AVIMYTSVTPNATIHKNKNTATWGGG-----KTVDQKTSYSGDVTVKYTI 204  
 DB 173 NVGEKTPSPTEIPFKYG-----WVDANNPSLIHWVRVNYAQNIPNAVFTDIIIGAKOTL 228  
 QY 205 TYKNAVNVHGTGKVKYQYVVKIDTPMSASVVDLNEGSYEVTITDGSNITTLTGGSEKATGK 264  
 DB 229 NFDISKAFHGTYSTDRVTAGTPIENTPFSKTSDFSVTL-----GNLDSVQIS----- 278  
 QY 265 YNLBENNFTTIPWAAATNTPTGNTQNGANDFFYKGINITVTYTGVLKSGAKPGSGAD 324  
 DB 279 -----YTTATDGGKSTQYDNTAKLSGDEFTVKOTSTWTPASGG---GGGADGTT- 325  
 QY 325 LPENTNIATINPNTSDPGQKVTYRDGQITIKIKIDGSTKASLQGAIFVLKNAATGQFLNF 384  
 DB 326 -----GSVTLTKQDAKTATLEGAEPKLVDAANGTILQES 358  
 QY 385 NDTNNVWTEANATEYTTGADGIITGLKEGTYLVLEKAPLGVNLLDNSOKVILGDG 444  
 DB 359 N-----ITDASGLNVSNLKPATYQLIETKAPTGYKLDSTPVEFTIGEN 403  
 QY 445 ATD-TTNSDNLVNPVTENK 464  
 DB 404 NKEITVTKENTLDTGAVELTK 424

## RESULT 25

Q8CYI8 PRELIMINARY; PRT; 2551 AA.  
 AC Q8CYI8;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN SPRI403.  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
 DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
 McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,  
 Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,  
 Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,  
 Glass J.I.;  
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
 RL J. Bacteriol. 183:5709-5717(2001).  
 DR EMBL: AE008509; AAL00207.1; --  
 DR GI: B98047; B98047.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR InterPro: IPR008161; Clg\_helix.  
 DR InterPro: IPR008160; Collagen.  
 DR InterPro: IPR001298; Filamin.  
 DR InterPro: IPR005877; Gpos\_Y5IRK.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR Pfam: PF01391; Collagen; 10.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam: PF04650; Y5IRK\_signal; 1.  
 DR ProDom: PD000007; Clg\_helix; 1.  
 DR TIGRFAMs: TIGR01167; LPXTG\_anchor; 1.  
 DR TIGRFAMs: TIGR01168; Y5IRK\_signal; 1.  
 DR PROSITE: PS0194; FILAMIN\_REPEAT; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 2551 AA; 265126 MW; B3A55D0AE30DA146 CRC64;

Query Match 8.2%; Score 212; DB 16; Length 2551;  
 Best Local Similarity 23.3%; Pred. No. 0.19;

Matches 121; Conservative 57; Mismatches 194; Indels 148; Gaps 24;  
 QY 34 TITVQDTQKATYKAYKVFDAIDN-----ANVSDSNKDGASYLI 73  
 DB 1891 TFTVKGKNGKDGKPIKIKVEDITSPIRRDTRDAAATPRNGIRVTVVDDVNDNGV-- 1948  
 QY 74 PQKRAEYKASTDFNSL-----FTTTNGGTYTKDGTASANEIATWAKSISANT 124  
 DB 1949 DEGVDKLVNKSIDYNGIDGRGSAFTITTKONGDGTHTITVQNPDSSTTVVKGKQK 2008  
 QY 125 TPVSTVTSNNNDGTREVINVSQYGYVSVSTVNGGAVIMVTSVTPNATIHKN-----TDA 179  
 DB 2009 TANITTE-NPDGSHITVTPNPDGSKETVVGKNGK-----DGKTPKVEVTDNDGTHTVKV 2063  
 QY 180 TWGDCG-----GKTVDQKTSYSGDVTVKYTYTITKNAVNVHGTGKVKYQYVVKID 225  
 DB 2064 TDGDNVTNNAIKGKQKGAATATTENPDG-SHTVITNP---DGTKN---EFVVGKGRD 2117  
 QY 226 ----TTPMSASVVDLNEGSYEVTITDGSNITTLTGGSEKATGKYNLLENNFTTIPWA 281  
 DB 2118 GVDGRTPTASVRDNGDGSHTIVITNPEG-VTTETTVRDGSKPKVTITDB----- 2165  
 QY 282 ATNTPGTQNGANDDFYKGINITVTYTGVLKSGAKPGSADLPENTNIATINPNTSD 341  
 DB 2166 -----QNGTHKISVLNGDGTITET---IIKQKSPVATVRDNDGTYTIRVENGNG 2213  
 QY 342 DPGQKTVVRDQ-----ITIKIDGSTKASLQGAIFVLKNAATGQFLNFND 386  
 DB 2214 TVSE-TTVRDGKSPAKVVDNGDGTHTITVNSDGIITTT-----TVRDGKPKLEVID 2266  
 QY 387 TNNVWTEANATEYTTGADGIITGLKEGTYLVLEKAPLGVNLLDNSQ----- 437  
 DB 2267 NN-----DGSHTIKVTGADGKGTITTFDG-----KSPKA-NIVDNGDGTHTLTIVD 2312  
 QY 438 -----KVILGDGATDTTNSDNLVNPVT---ENKGTTEL 468  
 DB 2313 SDGREYKSIIDKG-KDGKDS-----VSPVTYVKNNDGTHV 2347  
 RESULT 26  
 Q836L9 PRELIMINARY; PRT; 476 AA.  
 ID Q836L9  
 AC Q836L9;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Cell wall surface anchor family protein.  
 GN EF1092  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=V583 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927;  
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,  
 Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,  
 Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,  
 Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,  
 Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,  
 Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 Enterococcus faecalis.";  
 RL Science 299:2071-2074(2003).  
 DR EMBL: AE016950; AAO80892.1; --  
 DR TIGR: EF1092;  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR PROSITE: PS50847; GRAM\_POS\_ANCHORING; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 476 AA; 53355 MW; 1D957EDD3009D99D CRC64;

Query Match 8.1%; Score 208.5; DB 16; Length 476;



```
QY 125 TPVS-TVTESNDGTEVINVSQYGYVSSSTVNNGAVI-----MVTSTVTPNAT----- 171
DB 143 TPEVALNPTNENGVTNNSAAAAGLYLIRETSTPAGVIPAQDFLLAVPMTNPEGTGWLT 202
QY 172 --IHEKNTDATWGDGGKTKVQ- KTVSVGDTVKYTYITYKNAVNYHGTBKVYQYVTKDTP 228
DB 203 IYVYPKGSFTS-----ASKSVDNAASLKTDGPTWTIE-----AGIPLIRSHTSQDPVA 251
QY 229 SA--SVVDLNEGSYEVTIDSGNITTLTQSEKATGKNLLENNNFTITIPWAATNTP 286
DB 252 PAEFKIVD--TENTADLELVGGANGVTVPVTEFTVADY-----TVDV----- 293
QY 287 TQNTONGANDDFYKGINITVITY-----TGVLKSGAKPGSADLPE 327
DB 294 -----NPVVD-----GLTTVTITFNEGGKLVGKDDVTVELKTTVLKAG-----E 334
QY 328 NTNIAITINENTNDPPGQ-----KVTVRDGOITIKKIDGSTKASIQAGAFVLKKNATGOF 381
DB 335 INTTATI---TATDDTRATVTVDIKQTVKYGVTLTK-----KNQNGQL 375
QY 382 LNFNDNNVWETEANA-----KVTVRDGOITIKKIDGSTKASIQAGAFVLKKNATGOF 408
DB 376 A--KDAVFRVYATEAHAKAGNDHLITATNTGKWTGGKVLDDGFRFSDFADGADQV 433
QY 409 IITGLKECTYVYVEKAPLGNLNDNSOKVILGDGATDTTNSDNLNVPYVNNKGTTEL 468
DB 434 KGADGSLYQTYWLVENTAPAGQQL--AEPVFP-----TVTEVEDTLEVNTA--NTNAFVL 486
QY 469 PSTGGIGTTFIYIGAILVIGAGIVLVARRR 499
DB 487 PUTGGTGTAMLTILG--IGILAIVLFVARSR 515

RESULT 29
Q929J4
ID Q929J4 PRELIMINARY; PRT: 1622 AA.
AC Q929J4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Probable cell surface protein (LPXTG motif).
GN LIN2281.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,
RA Entian K.-D., Fslhi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerel U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596171; CAC97509.1; --
DR PIR; AE1717; AE1717.
DR ListiList; LIN02281; --
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR008456; Collagen bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 6.
DR Pfam; PF05737; Collagen_bind; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAM; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; P850847; GRAM_POS_ANCHORING; 1.
```

```
KW Complete proteome.
SQ SEQUENCE 1622 AA; 178366 MW; 7A12A5775BA6FF37 CRC64;

Query Match      8.0%; Score 205; DB 16; Length 1622;
Best Local Similarity 22.4%; Pred. No. 0.23;
Matches 134; Conservative 65; Mismatches 207; Indels 192; Gaps 27;

QY 18 MAVSPVPIAFA--AETGTTIV-----QDTQK---GATYKAYKVFDAIDNANVSDSNK 66
DB 991 VAVSPVFTVGTGSGIATIGTIKITKVDDEDTTKLEGAKFQLY-TLDGEKSGQEIT-TNS 1048
QY 67 DGASVYL--IPOGK-----EAETK-----ASTDFNSLFTTTNG---GRTY 101
DB 1049 EGEILLDGIQSGKYLKVTKAPEGVNIIDEVKEGKEITVSSSGEELLITIKAMKKGVI 1108
QY 102 VTCKDTASANI-----ATWAKSISANTPVSSTVTSNN----- 135
DB 1109 LTKDSASHEVLADAEFELQNTAGSKLEKLTDDASGNIETDLAPGDYKLIETKAPTGY 1168
QY 136 --DGTEV-----INVSYQYVYVSVTVNGAVIM-----VTSVTPNATIIHKNTDATWG 182
DB 1169 QLDATFPVNTIDFNGQSEAAKVSKTNTAKTGTVMLTCKDSATNAELADATFELRNEDETLV 1228
QY 183 DGGKTVDPQKTVSYGD-----TVKYTYTKNA-----VNVHGT 216
DB 1229 RENLVTDDNGEISVADLAPGDYKLIETKAPTGYQLDAPVNFITIDFNGQSEVSKVTNTA 1288
QY 217 KVVQYVI--KDTMPSASVVDL-----NEGS---YEVITIDGSGNITTLTQSEKATGKN 266
DB 1289 KGTGTVLTCKDSATNAELADATFELRNESDILVRENLTDDNGEISV---ADLAPGDYK 1344
QY 267 LLENNNFTITIPWAATNPTGNTONGANDDFYKGINITVITYTVTVGLKSGAKPGSADLP 326
DB 1345 LIE-----TKAPAGYQLDAPVNF----- 1363
QY 327 ENTNIATINPNTSNDPPQKV--TVRDGOITIKKIDGSTKASI-QGAFVLKKNATGQFLNFN 385
DB 1364 -----TIDFNGQSEAAKVSKTNTAKTGTVLTCKDSATNAELADATFELRNESDILVREN 1417
QY 386 DTNNVWGTGANATEYTTGADGIITGLKEGYVYLVKPKAPLGYNL-----LDNSQ 437
DB 1418 -----LVTDNGEISVADLAPGDYKLIETKAPTGYQLDAPVNFIDFNQ 1462
QY 438 KVLGDGATDTTNSDNLVN-PTVNNK-GTELPSTGGIGTTFIYIGAILVIGAGIV 493
DB 1463 SEAAKVTKNKKKIGTIIVNFIDVGNQNDKEVHTGNVGE--YSVKAKEIVGHTLV 1518

RESULT 30
Q7WN54
ID Q7WN54 PRELIMINARY; PRT: 3346 AA.
AC Q7WN54;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative hemolysin.
GN BB1186.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=2827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achman M., Atkin R., Basham S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Gobie A., Hamlin N., Hauser H., Holroyd S., Jagsis K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond B., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 10:26:31 ; Search time 19 Seconds  
(without alignments)  
1364.012 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKQMIQSLVSLAFGMV.....GAILVIGAGIVLARRRLRS 502

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/iaa/PCITUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334	13.0	627	4	US-09-071-035-218
2	334	13.0	659	4	US-09-134-000C-6124
3	318	12.3	664	4	US-09-107-532A-7252
4	274.5	10.6	560	4	US-09-071-035-220
5	234.5	9.1	341	4	US-09-107-532A-5384
6	219	8.5	915	4	US-09-206-942-35
7	219	8.5	1222	4	US-09-206-942-37
8	219	8.5	1228	4	US-09-206-942-34
9	208.5	8.1	477	4	US-09-134-000C-6123
10	208.5	8.1	660	4	US-09-107-532A-6715
11	205.5	8.0	1027	4	US-09-107-532A-6675
12	199	7.7	508	4	US-09-107-532A-5331
13	194	7.5	1638	4	US-09-071-035-258
14	194	7.5	1638	4	US-09-071-035-222
15	194	7.5	1638	4	US-09-071-035-286
16	192.5	7.5	1095	4	US-09-206-942-43
17	192.5	7.5	1101	4	US-09-206-942-45
18	192	7.4	1220	4	US-09-206-942-28
19	192	7.4	1226	4	US-09-206-942-26
20	192	7.4	1747	4	US-09-134-000C-5999
21	190.5	7.4	1073	4	US-09-206-942-49
22	190.5	7.4	1079	4	US-09-206-942-47
23	190	7.4	669	4	US-09-071-035-264
24	186.5	7.2	1742	4	US-09-386-962C-4
25	185.5	7.2	1026	1	US-08-194-290-7
26	183.5	7.1	2736	4	US-09-252-991A-30227
27	183	7.1	2411	4	US-09-268-347-36

#### ALIGNMENTS

RESULT 1

US-09-071-035-218

; Sequence 218, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

28	183	7.1	3892	4	US-09-328-352-5503	Sequence 5503, Ap
29	181.5	7.0	1315	4	US-09-200-650E-5	Sequence 5, Appli
30	181.5	7.0	1833	4	US-08-621-944A-4	Sequence 4, Appli
31	181.5	7.0	1833	4	US-08-945-567D-4	Sequence 4, Appli
32	181.5	7.0	1992	4	US-08-621-944A-3	Sequence 3, Appli
33	181.5	7.0	1992	4	US-08-945-567D-3	Sequence 3, Appli
34	181.5	7.0	2048	4	US-08-268-347-48	Sequence 48, Appli
35	181	7.0	673	4	US-09-107-532A-5134	Sequence 5134, Ap
36	180.5	7.0	413	4	US-09-107-532A-6433	Sequence 6433, Ap
37	180.5	7.0	1180	4	US-09-206-942-65	Sequence 65, Appl
38	180.5	7.0	1188	4	US-09-206-942-63	Sequence 63, Appl
39	180	7.0	1166	4	US-09-200-650E-7	Sequence 7, Appli
40	180	7.0	1231	3	US-08-904-283A-4	Sequence 4, Appli
41	180	7.0	1231	4	US-09-434-123A-4	Sequence 4, Appli
42	180	7.0	2353	3	US-09-377-155-33	Sequence 33, Appl
43	180	7.0	2353	3	US-08-913-942-4	Sequence 4, Appli
44	180	7.0	2353	4	US-09-669-974-33	Sequence 33, Appl
45	180	7.0	2353	4	US-09-797-862-33	Sequence 33, Appl
46	180	7.0	2354	4	US-09-268-347-47	Sequence 47, Appl
47	179.5	7.0	905	4	US-09-134-000C-4420	Sequence 4420, Ap
48	177.5	6.9	1183	2	US-08-447-031A-2	Sequence 2, Appli
49	177	6.9	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
50	175.5	6.8	1536	1	US-08-038-682-2	Sequence 2, Appli
51	175.5	6.8	1536	1	US-08-302-832-2	Sequence 2, Appli
52	175.5	6.8	1536	2	US-08-530-138-2	Sequence 2, Appli
53	175.5	6.8	1536	2	US-08-469-880-2	Sequence 2, Appli
54	175.5	6.8	1536	2	US-08-728-470-2	Sequence 2, Appli
55	175.5	6.8	1536	2	US-08-617-697-2	Sequence 2, Appli
56	175.5	6.8	1536	3	US-08-719-641-2	Sequence 2, Appli
57	175	6.8	1095	4	US-09-206-942-69	Sequence 69, Appl
58	175	6.8	1536	4	US-09-206-942-67	Sequence 67, Appl
59	175	6.8	2314	4	US-09-268-347-49	Sequence 49, Appl
60	174	6.7	1026	3	US-08-614-377A-7	Sequence 7, Appli
61	174	6.7	1026	3	US-09-142-648B-7	Sequence 7, Appli
62	174	6.7	1233	4	US-09-134-000C-4971	Sequence 4971, Ap
63	174	6.7	1912	1	US-08-409-995-4	Sequence 4, Appli
64	174	6.7	1912	3	US-08-685-467-4	Sequence 4, Appli
65	173.5	6.7	952	4	US-09-107-532A-4706	Sequence 4706, Ap
66	173	6.7	1004	4	US-09-268-347-30	Sequence 30, Appl
67	173	6.7	1338	2	US-08-728-470-9	Sequence 9, Appli
68	173	6.7	1338	3	US-08-719-641-9	Sequence 9, Appli
69	173	6.7	1416	4	US-09-071-035-404	Sequence 404, App
70	173	6.7	1448	4	US-09-071-035-402	Sequence 402, App
71	173	6.7	1529	2	US-08-728-470-10	Sequence 10, Appl
72	173	6.7	1529	3	US-08-719-641-10	Sequence 10, Appl
73	173	6.7	1599	2	US-08-617-697-9	Sequence 9, Appli
74	173	6.7	1600	2	US-08-617-697-10	Sequence 10, Appl
75	172.5	6.7	1104	4	US-09-268-347-28	Sequence 28, Appl
76	172.5	6.7	1104	4	US-09-268-347-34	Sequence 34, Appl
77	172.5	6.7	2032	4	US-09-071-035-458	Sequence 458, App
78	172.5	6.7	2032	4	US-09-071-035-462	Sequence 462, App
79	172.5	6.7	2032	4	US-09-071-035-466	Sequence 466, App
80	172.5	6.7	2054	4	US-09-134-000C-6612	Sequence 6612, Ap
81	171	6.6	2142	4	US-09-540-236-3459	Sequence 3459, Ap
82	169	6.6	2504	4	US-09-328-352-5821	Sequence 5821, Ap
83	168	6.5	1132	4	US-09-198-452A-4861	Sequence 4861, Ap
84	167	6.5	507	4	US-09-134-000C-4961	Sequence 4961, Ap
85	166	6.4	1612	1	US-08-169-927-2	Sequence 2, Appli
86	164.5	6.4	558	4	US-09-071-035-268	Sequence 268, App
87	164	6.4	1098	1	US-08-409-995-2	Sequence 2, Appli
88	164	6.4	1098	3	US-08-685-467-2	Sequence 2, Appli
89	164	6.4	1098	3	US-09-377-155-32	Sequence 32, Appl
90	164	6.4	1098	3	US-08-913-942-2	Sequence 2, Appli



Db 627 STGGKGIYVYLGSGAVLLLIAGVYFARRK 656

US-09-107-532A-7252

RESULT 3

Sequence 7252, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 7252:

SEQUENCE CHARACTERISTICS:

LENGTH: 664 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURES:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...664

SEQUENCE DESCRIPTION: SEQ ID NO: 7252:

US-09-107-532A-7252

Query Match 12.3%; Score 318; DB 4; Length 664;

Best Local Similarity 24.2%; Pred. No. 2.1e-16;

Matches 141; Conservative 91; Mismatches 206; Indels 144; Gaps 28;

QY 3 KKMIGSLVSLAFNGMAVSPVTPIAFAA--ETGTTV-----QDTOKGATYKAYKVFDAEI 56

Db 137 KQAVQSL-----TPGTPVAGSTDDAGNVTLSPKKQNGDAVY----- 175

QY 57 DNANVSDSNKGAS-----YLI PQKEAEYKAST---DFNSLF--TTTNGGRT 100

Db 176 ---TIKEEPKDGVSAAANNVLAFFVYEMIKQADGSYKGTBELDTHLYPKTVGNDGTL 232

QY 101 YVTKDGTASANEIATWAKSISANT--TP-----VSTVTE----- 132

Db 233 KVTIKGTAE-NEALANGAEFIISKEGTPSVKVIQSVTDGLYTTTDDTKAKHFITGHSY 291

QY 133 --SNND-----GTEVINVSQGYGVYVS--STVNGAVIMVTSVTPNATHEKNT--- 177

Db 292 DIGNNDFAEASIEKGLQILVHLEVGKYNLEEVKAPDNAEMIEKQTIITPPEILANSQTPVE 351

QY 178 -----DATMGDGGKTVDOQKTYTSVGDVTVKYITTK---NAVNYHGTEKY-QYVIKOTMP 228

Db 352 KTIKNDTSKVDKTTPLQNGKDVAIKEKIQEISVNIPLGIADKSGTQNKYTTFKLIDTHD 411

QY 229 SASVVDL-NEGSYEVITIDGSGNITLTQSGSEKATGKYNLLENNNFITIP--WAAANT 285

Db 412 AALTFDNDSSGTAYALYDGNKEIDPV-----NYSVTQDTGFTVSDPNTIPSLT 462

QY 286 PTGNTQNGANDFFYKGINITITVYTVGLKSGAKPGSADLPENTNIATINPNTSNDPQ 345

Db 463 PGG-----TLKFVYMHLEKADPTKG----FSNQANVDNGHTNDQTTP 502

QY 346 KTVRDGQITTKIDG--STKASIQGALFVLKNA---TGQFLNFN-DTNVVEW-GTEANA 398

Db 503 SVDVVVGKRFVKVDGVDTSQTLAGAEFVVVRDQSDSTAKYLSIDPSTKAVSWSAKESA 562

QY 399 TEYTTGADGIITITLKGEGTYLVEKKAPLGNLLDNSQKVLGIDGATDNTNSDNLVNP 458

Db 563 TVFTTTSNGLIDVTGLKYGTYLLETKAPEKVPVLTNRVAFITDEQSVYTAGQ----LISP 619

QY 459 --TVENNKGTSLPSTGGIGTTFIYIIGAILVIGAGIVLVARR 498

Db 620 EKIPNKHGT-LPSTGGKGIYVYIGAGVVLIIAGLYFARRK 660

RESULT 4

US-09-071-035-220

Sequence 220, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 220:

SEQUENCE CHARACTERISTICS:

LENGTH: 560 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-071-035-220

Query Match 10.6%; Score 274.5; DB 4; Length 560;

Best Local Similarity 25.3%; Pred. No. 3.9e-13;

Matches 136; Conservative 72; Mismatches 212; Indels 117; Gaps 25;





Db 309 AKDNSNLITGDSNDAKGV-----TFSNVKDSKISASDHNVTLSKVETSGDTST 363  
QY 87 -----FNSLFTTTTNGRVTYTKDOTASAN-----E1ATWAKSISANTTP 126  
Db 364 EDGNNNTGLTITAKNVTVNNITSHKTVNITASENVTTKAGTTINATTSVEVTAKTGD 423  
QY 127 VSTVTESSNDGTEV-----INVSQGYGYYVSSVTNNGAVIMVTSVTPNATHEKNTDAT 180  
Db 424 IKGGIESGNVNITASGDTLNVSNITQONVTVAASGAVTTTKGSTINATNGNANITTK 483  
QY 181 WDGGGKTVQDKTYSVGDTVK-YTITYKNAVNHGTEKYQYVVIKDTMPSASVDLNEGS 239  
Db 484 TGEINGE-VKSAGNVNITASGDTLNVSNITQONVTVTANSgai--TTEGSTINATNGD 540  
QY 240 YEVITIDG-----SGNITTLTQSEKATGKYNLLENNNFTITIPMAATNPTGNT 290  
Db 541 ANITQTQNGINGKVESGSSVTLIATGOTLAVGNIS-----GDTVTITADKGLTTOTSSK 596  
QY 291 QNGANDDPFYKGI-----NTITVTYTVGL-----KSGAKPGSADLPENT 329  
Db 597 INGT-----KSVTSSQSGDISGTVSVSATGSLTTQAGSKIEAKTGEANVTSAT 650  
QY 330 NI--ATINPNTSNDPPQKVTVRDGOITIKKIDGSTKASLOGAI FVLKNATGQFLNFNDT 387  
Db 651 GTIGTISGTVN-----VTANTDNLTIK--DGARIKATGGAVTL--TATGGLTTTETS 700  
QY 388 NNVEWG-----TEANATEYTTGADGIITITGLK-----EGTYLVEKKA 426  
Db 701 SDITSSNGQTLTAKDSSIFAGSINAANVTLTGT--LTTVAGSKIEAASGTLVINAKDA 758  
QY 427 PLGYNLLDNSOKVILGDGATDTNSDNLVNPTVENNKGTELPSGTGGITTFIFYIGAIL 486  
Db 759 QL-----DGA---ASGDHTVNVNATNANGSGSVIATT-----SSRVNITGDLI 797  
QY 487 VIGAGIVLVAR 497  
Db 798 TIN-GLNIISK 807

RESULT 7  
US-09-206-942-37  
; Sequence 37, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 37  
; LENGTH: 1222  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-37

Query Match 8.5%; Score 219; DB 4; Length 1222;  
Best Local Similarity 22.5%; Pred. No. 2.5e-08;  
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;  
QY 29 AAGTGTITVQDTQKATYKAYKVFDAEIDNANVSNSKDGASYLIPQKGEABYKASTD-- 86  
Db 616 AKDNSNLITGDSNDAKGV-----TFSNVKDSKISASDHNVTLSKVETSGDTST 670  
QY 87 -----FNSLFTTTTNGRVTYTKDOTASAN-----E1ATWAKSISANTTP 126  
Db 671 EDGNNNTGLTITAKNVTVNNITSHKTVNITASENVTTKAGTTINATTSVEVTAKTGD 730

QY 127 VSTVTESSNDGTEV-----INVSQGYGYYVSSVTNNGAVIMVTSVTPNATHEKNTDAT 180  
Db 731 IKGGIESGNVNITASGDTLNVSNITQONVTVAASGAVTTTKGSTINATNGNANITTK 790  
QY 181 WDGGGKTVQDKTYSVGDTVK-YTITYKNAVNHGTEKYQYVVIKDTMPSASVDLNEGS 239  
Db 791 TGEINGE-VKSAGNVNITASGDTLNVSNITQONVTVTANSgai--TTEGSTINATNGD 847  
QY 240 YEVITIDG-----SGNITTLTQSEKATGKYNLLENNNFTITIPMAATNPTGNT 290  
Db 848 ANITQTQNGINGKVESGSSVTLIATGOTLAVGNIS-----GDTVTITADKGLTTOTSSK 903  
QY 291 QNGANDDPFYKGI-----NTITVTYTVGL-----KSGAKPGSADLPENT 329  
Db 904 INGT-----KSVTSSQSGDISGTVSVSATGSLTTQAGSKIEAKTGEANVTSAT 957  
QY 330 NI--ATINPNTSNDPPQKVTVRDGOITIKKIDGSTKASLOGAI FVLKNATGQFLNFNDT 387  
Db 958 GTIGTISGTVN-----VTANTDNLTIK--DGARIKATGGAVTL--TATGGLTTTETS 1007  
QY 388 NNVEWG-----TEANATEYTTGADGIITITGLK-----EGTYLVEKKA 426  
Db 1008 SDITSSNGQTLTAKDSSIFAGSINAANVTLTGT--LTTVAGSKIEAASGTLVINAKDA 1065  
QY 427 PLGYNLLDNSOKVILGDGATDTNSDNLVNPTVENNKGTELPSGTGGITTFIFYIGAIL 486  
Db 1066 QL-----DGA---ASGDHTVNVNATNANGSGSVIATT-----SSRVNITGDLI 1104  
QY 487 VIGAGIVLVAR 497  
Db 1105 TIN-GLNIISK 1114

RESULT 8  
US-09-206-942-34  
; Sequence 34, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-34

Query Match 8.5%; Score 219; DB 4; Length 1228;  
Best Local Similarity 22.5%; Pred. No. 2.5e-08;  
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;  
QY 29 AAGTGTITVQDTQKATYKAYKVFDAEIDNANVSNSKDGASYLIPQKGEABYKASTD-- 86  
Db 622 AKDNSNLITGDSNDAKGV-----TFSNVKDSKISASDHNVTLSKVETSGDTST 676  
QY 87 -----FNSLFTTTTNGRVTYTKDOTASAN-----E1ATWAKSISANTTP 126  
Db 677 EDGNNNTGLTITAKNVTVNNITSHKTVNITASENVTTKAGTTINATTSVEVTAKTGD 736  
QY 127 VSTVTESSNDGTEV-----INVSQGYGYYVSSVTNNGAVIMVTSVTPNATHEKNTDAT 180  
Db 737 IKGGIESGNVNITASGDTLNVSNITQONVTVAASGAVTTTKGSTINATNGNANITTK 796

QY 181 WGGGKTVQDKTYSVGDVTK-YTITYKXVNVHGTKEKYQVVIKDTMPSASVVDLNEGS 239  
Db 797 TGEINGE-VKSASGNVNIITASGNLTNVSNIQGVTVTANSAGAI--TTTEGSTINATTGD 853  
QY 240 YEVITIDG-----SGNITTLTQSEKATGKYNLLENNFTIIPWAAATNPTGNT 290  
Db 854 ANITTTQGNINGKVESSSGVLIATQOTLAVGNIS-----GDTVIITADKGLTKTQTSK 909  
QY 291 QNGANDDFYKGI-----NNTIVYTVGVL-----KSGAKPGSADLPENT 329  
Db 910 INGT-----KSVTTSSQSGDISGTTISGNTVSVSATSLTQAGSKIEAKTGEANVTSAT 963  
QY 330 NI--ATINPNTNDPGQVTVRQDGOITIKKIDGSTKASLOQAI FVLKNATGQFNFNDT 387  
Db 964 GTTGTISGNTVN-----VTANTDNLTK--DGARIKATGGAVTL--TATGTTLTETS 1013  
QY 388 NNVEWG-----TEANATEYTTGADGIITITGLK-----EGTYVLVEKKA 426  
Db 1014 SDITSSGQTLTAKDSSTAGSINAANVLTNTGT--LTTVAGSKIEAASGTLVINAKDA 1071  
QY 427 PLYNLLDNSQKVLGQATDTTNSNLLVNPVVENKGTLEPSTGGITTFYIIGAIL 486  
Db 1072 QL-----DGA---ASGHTVVNATNANGSGSVIATT-----SSRVNITGDLI 1110  
QY 487 VIGAGIVLVAR 497  
Db 1111 TIN-GLNIISK 1120

RESULT 9  
US-09-134-000C-6123  
; Sequence 6123, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6123  
; LENGTH: 477  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-6123

Query Match 8.1%; Score 208.5; DB 4; Length 477;  
Best Local Similarity 26.5%; Pred. No. 4.1e-08;  
Matches 85; Conservative 49; Mismatches 124; Indels 63; Gaps 14;  
QY 187 KTV--DQKTVSVGTIVYITYKXVNVHGTKEKYQVVIKDTMPSASVVDLNEGSYEVTI 244  
Db 193 KTVLDKQGGQNGEHNLYQLTQIPANILG---YQ-----EFRL 228  
QY 245 TDGSGNITTLTQGS--EKATGK-----YNLLENNFTIIPWAAATNPTGNTQNGANDD 297  
Db 229 SDKADTTLTLTPESIEVKVAGKVTVTGYTLTQKHGFTL-----DFSIKDLQNFAN-- 279  
QY 298 PFYKGINITVYTVGLVSKGAKPGSADLPENTINATINPNTSDPGQKTVTRDGOITIK 357  
Db 280 -----QTMVTSYQWRLEKTAEPDTAINNEQLV-----TDKHTLTKRATVTRGGKSFV 327  
QY 358 KIDG-STKASLQGAIFVLKNATGQFNLNDNNVWGTGTE-ANATEYTTGADGIITITGLK 415  
Db 328 KVDSENAKITLPEAVFVVKNOAGEYLN--ETANGYRWQKEKALAKKFTSNQAGEFSVKGLK 386  
QY 416 EGTYYLVEKAPLGVNLLDNSQKVLGQATDTTNSNLLVNPVVENKGTGTE---LPSTG 472  
Db 387 DQOIFLEISAPKGYLLNQTEIPFTVGNKSTATNQRTAPLH--VINKVKESGFLPKTN 444

QY 473 GIGTTFIICAILVIGAGIV 493  
Db 445 E-ERSIMLTIAGLLIIGMVVI 464  
RESULT 10  
US-09-107-532A-6715  
; Sequence 6715, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arinello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 6715:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 660 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...660  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6715:  
US-09-107-532A-6715  
Query Match 8.1%; Score 208.5; DB 4; Length 660;  
Best Local Similarity 23.4%; Pred. No. 6.5e-08;  
Matches 127; Conservative 62; Mismatches 190; Indels 163; Gaps 25;  
QY 10 LVASLAF-----GMVSPVTPIAFAAETGTTITVQDTQKATYKAYKVFAEIDNANV 61  
Db 210 IVSKLSFEUTKFGDDGTAQSKQTPLE-----KGAKFLYK-----243  
QY 62 SDSNKGDSYLIIPQKGEABYKASTDFNSLFTTTTNGRTYVT-----KKDTASA 110  
Db 244 ---GEPG-----KGIKGLDVLSDQKGLTADTLGLKYFVEVPSEVVVSGSKPTADQ 294  
QY 111 NEATWAKSISANTPTPVSTVTSNNDG--TEVINVSQYGY--YVSSTVANGAVIMVTSVT 167  
Db 295 YLLGADARNDAHN-----KLTFEITNDGVTSDLKASTVYKAPVIDKTVTNG-----341

```
QY 168 PNATHEKNTDATWGDGGKTVQDKTYSYVGDIVKYITTYKNAVNYHTEKYVQYVTKDM 227
Db 342 -----TGOEHSFQIGD-----AVNYQGT-----IHIPTDIA 367
QY 228 PSASVVDLNEGSYEVT-----ITDGSNNITLTQSEKATGKYN--LLENNNNFTIT 277
Db 368 CGADGITVNGVKSETSPYSVFKWGTAGQGLSVAAKANIKVTKDGSVVLKENTDYKIQ 427
QY 278 IPWAATNTPTGNTQNGANDPFYKGINITVTVYTVGLVKSAGK-PSGADLPENTNNTATNP 336
Db 428 -----NSENGVIDFI---VNNQVSDTVASLHGQDLQMTYNNVYVNSAAVAMP 473
QY 337 NTS-----NDDP-----GQKTVRDGQITIKID-----GSTKASLOGAIFVLKN 376
Db 474 LTNVDVFNPNPFQOEHEHETKADVTVYGAFLKLVDSGLFGTGIGATPLESABFAKN 533
QY 377 ATGQF---LNFNDTNVVE---WGTE-ANATEYTTGADGIITITGLKEGYIYLVKKAPLG 429
Db 534 ASGKYGGLVDTDKNGKVAVVDVANAAILKSDKEGHFEITGLTEGEYSLEETKAPEN 593
QY 430 YNLLNSQKVLGSGATDTNSDNLVNPVTNNKGTLPSTGGIGTTIFYIIGAILVIG 489
Db 594 YQKL---TKEISFKVDKDSYKEENRI---TIKNQKASVPMTGSGNGFQYVVLISCLL-LG 646
QY 490 AG 491
Db 647 AG 648

RESULT 11
US-09-107-532A-6675
; Sequence 6675, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6675:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1027 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
```

```
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1027
; SEQUENCE DESCRIPTION: SEQ ID NO: 6675:
US-09-107-532A-6675

Query Match      8.0%; Score 205.5; DB 4; Length 1027;
Best Local Similarity 21.7%; Pred. No. 2.1e-07;
Matches 127; Conservative 67; Mismatches 179; Indels 211; Gaps 26;

QY 23 VTPIFAAETGTTIVQ--DTOKGATY--KAYKVPDAEIDNANVSDSNKDGASYLIPQOK 77
Db 424 VDKVSFNKETNLVVDVFGNLAEGQSYFIEYAUEVTDVDFN--EGROEDGFS----- 473
QY 78 EAEEKASTDFNSLFTTTTNGGRTYVTK-----KDTASANE-IATWAKSI 120
Db 474 --RNRASMTYNNQVMS-----RDYELRIPESNLPALSRKSVDKVINLENELIYTLHV 526
QY 121 S-----ANTPTVSTVTBSNNDGTEVINVSQYGYVVSSTVNNGAVIMVTSVTPNATIHEK 175
Db 527 SMFSGTRNTLVIEDVLE--EGMEFVSFLSDIPSFIDAEVNG--IVRFSANNTI--- 576
QY 176 NTDATMGDGGGKT-----VDQKTSYSGDITVKYITTYKNAVNYHTEKYVQYVTKDTP 228
Db 577 -----GGPTSFTLSFKVNVEAYVGDQILNLFATLVNSDNLNTSTVTRTKI----- 622
QY 229 SASVVDLNEGSYEVTITDGSNNITLTQSEKATGKYNLLE-----ENNNTF-T 275
Db 623 -----DGRKIVKTD-----ENGQRLAGAKFEILSSDSQVIOEGITAEENGEPFLS 666
QY 276 ITIPWAA-----TNPNTGNTQNGANDPFYK--GINTITVTVYTVGLKSGA-----KPGS 322
Db 667 EPLPVGYSQVRETEAPEGYLLDATVHDVFINESYIEEVTLTIENTILDTGGVELIKHAEGS 726
QY 323 ADL-----PENTNIATINPN-TSNDP-----PG----- 344
Db 727 GEVLQAGVFNQNRREGETLQGLTTGDEGKLAIDGLAPGAYQVLVETQAPIGYELDATPIE 786
QY 345 -----OKVTVRDQGITIKIDGSTKASLOGAIFVLKNATGQFLNFDNTNNV 390
Db 787 FEIERSQTAVVVELTKENRLTPGGVVLTKIDDSGILQAGVFNQNRREGETLQ----- 839
QY 391 EWGTEANATEYTTGADGIITITGLKEGYIYLVKKAPLGYNL-----LDNSQKVLIG 442
Db 840 -----TGLTTGDEGKLAIDGLAPGTYQLVETQAPIGYELDATPIEFIERSQTAV-- 889
QY 443 DGATDTTNSDNLVNPVTNNKGTLPSTGGIGTTIFYIIGAIL 486
Db 890 -----VELTKENRLTPGGVVLTKIDDSGSEIL 916

RESULT 12
US-09-107-532A-5331
; Sequence 5331, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GFC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 5331:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...508
; SEQUENCE DESCRIPTION: SEQ ID NO: 5331:
US-09-107-532A-5331

Query Match
Best Local Similarity 21.5%; Pred. No. 2.4e-07;
Matches 118; Conservative 84; Mismatches 216; Indels 130; Gaps 23;

QY 1 MKKMQISLVASLAEQMAVPTPTAFATETGTVQ-----DQ 41
Db 36 MKKLGLSMCLFLLLPKPAFTQV-----ATETEMVQIHLKLLPFGNOLPKRPNDDQ 90
QY 42 KGATYKAYKVFDAEIDNANVSDNKGASYLIPQKAEAYKASTDFNSLFTTTTNGRTV 101
Db 91 EKALLQYR-----GLNGVTFQV-----YDVTDSFVHL-----REKQTV 125
QY 102 VTKDPTASANEIATWAKSISANTPTVSTVTESNNDGTEVINVS-----QYGYVVSST 154
Db 126 -----EEAQA-EIAKNGASSGMFTAEATTTTNNEDGTASFSLAAKQKEDKAYLFIESK 180
QY 155 V-----NNGAVIMVTSVTPN---ATIH--EKNTDATWGDGGKTV---DQKTVSGD 198
Db 181 VPEVVEKAENMVVLPHVQNNQKUSTIHLXPNEENDYDPPPEKVLPEERNDFTGE 240
QY 199 TVKYTITYKNVNVHGTKEYQYQVVIKDTMPSASVVDLNEGSYEVVITDGSNNITTLTQGS 258
Db 241 KITYSLHTTTPNILDYQK---FELSDSADAE-----LTFLP---NSLTSSNG 283
QY 259 EKATGKNLLENNFTI--TIPWAATNPTGNTQNGANDDPFYKGINITVYTGVLKS 316
Db 284 EKLTEGVIIHKPHGFDVLPFSLEK-----YAG-KKLITISYQMLSS 326
QY 317 GAKPGSADLPENTNIATINPTNSD---DPGOKVTVRDQIITIKID-GSTRKASIQGAI 371
Db 327 TAQANK-----EINNGTLDGFGVSTKKVSVYTGSKQFVKLETNPKDKRLAGAV 376
QY 372 FVLKNATGQFLNFDNTNNVW-GTEANATPYTTGADGIITIGLKEGTYYLVEKAPLGY 430
Db 377 FLIKRAGNYLQ-QTANGYKWTKNESDALHLSDKNGAFSISGLKTSYRLKEIAPSGY 435
QY 431 NILDNSOKVILGQADTTNSDLLNVPVNNKGTSLPSTGIGITTFIIGAILVIGA 490
Db 436 ILSETEIPFTISTFLSEDKADSLKLVNKKNSRPFPLPKTNETKNTLLGVGVGMVFPASFA 495
QY 491 GIVLVARR 498
Db 496 IWLFIKKR 503
```

## RESULT 13

```

US-09-071-035-258
; Sequence 258, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 258:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-258
```

```

Query Match
Best Local Similarity 21.8%; Pred. No. 3.3e-06;
Matches 136; Conservative 56; Mismatches 219; Indels 212; Gaps 26;
```

```

QY 23 VTPIAFAETGTTTVDQTKGATYKA---YKVFDAEIDNANVSDNKGASYLIPQKSEA 79
Db 340 VSNIFYAAGTTSEVTAPTITGTGNSTAGYEVKGTADANATVEIRNAGGT--VIGTG--- 394
QY 80 EYKASTDFNSLFTTTTNGGRTVTKDPTASANEIATWAKSISANT-----TPVSVTV 131
Db 395 -----TADGTGAFTVTPAGE-----AGANETLTAVAKNASGTEXTPTTFTQTPADEAT 442
QY 132 ESNNDGTEVINVSQYGYVVSSTVANGAVIMVTSVTPNATHEKNTDAT-----WGDG-G 185
Db 443 VTAPTITGTGNSTAG-YEVKGTAD-----ANATVEIRNAGGTIVGTGTADGTG 490
QY 186 GKTVDQKTVSGDTPVKYTYITYKNVNVHGTKEYVQ-----YV 222
Db 491 AFTVTVPAGEAGANETLTAVAKNASGTESTPTTFTQTPADEATVTAPTITGTGNSTAGYE 550
QY 223 IKDWTMPSASVVDLNE-----GSEVETITDG-SGNITTLTGSEKATGKNYL 267
Db 551 VKGTADANATVEIRNAGGAVIGTGTADGTGAFTVTPAGEAGANETLTAVAKNASG----- 606
QY 268 LEENNNFTITIWAATNTP-----TGNTQNGANDDPFYKGINITITVT----- 309
Db 607 -TESTPTTFTQTP-ADPNTFPVATPIVETVGTGTTKG-----YEVKGTAEVGTITIEVRDAAGT 660
```

QY 310 -----YGVLSKSG-----AKPGSADLPENTNIATINPNT 338  
DB 661 VLGTATTGDKYTVTLDSGTATANTQTLVSVAKNAGSTESQATATTPADVTAPVDNIT 720  
QY 339 SNDDPGQKVT-VRDQGIITIKKIDGSTKASLOGAIF--VLKNATGQF-----LNFNDDT 387  
DB 721 GNSGSGYEITGTADPNNTTIEVRDPS-----GAVIGTGTSDANGDFTVTLPTGTINPGDT 774  
QY 388 NNVT-----EWGTEANATEYTTGADGIIT-----ITGLKEGTYVLVEKKAPLGLYNLLDNSQ 437  
DB 775 LTVIGKDNAGNESQPTVELVPADATVTAFTVGTGNSVAGYQVGTGADPNATIEIRDAD 834  
QY 438 KVLGDGATDTTNS-----NNKGTLPSTGGIGTTI 478  
DB 835 GNVIATGTADGTGSGFAVNLPAGTANANETLTALAKDPAGNTSTPTTFTQPADEVVAPPSV 894  
QY 461 E-----NNKGTLPSTGGIGTTI 478  
DB 895 DKVTGNTTQGYQVGTGAELGTTI 917  
  
RESULT 14  
US-09-071-035-262  
; Sequence 262, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 262:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1638 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-262  
  
Query Match 7.5%; Score 194; DB 4; Length 1638;  
Best Local Similarity 21.8%; Pred. No. 3.3e-06;  
Matches 136; Conservative 56; Mismatches 219; Indels 212; Gaps 26;  
  
QY 23 VTPIAFAETGTTVQDTQKGATYKA---YKVFDAEIDNANVSDSNKDGASLYLIPQKEA 79  
DB 340 VSNIFYAAGTSEVTAFTTGTGNSVAGYQVGTADANATVEIRNAGT--VIGTG--- 394  
QY 80 EYKASTDFNSLFTTTTNGGRTVTVTKDTASANEIATWAKSISANT-----TPVSTVT 131

DB 395 ----TADGTGAPTVTPAGE-----AGANETLTAVAKNASGTEXTPTTFTQPADEAT 442  
QY 132 ESNNDGTEVINVSQYGYVSVSTVNNGAVIMVTSVTPNATIEHKNTDAT-----WGDG-G 185  
DB 443 VTAPITITGTGNSVAGYQVGTAD-ANATVEIRNAGGTVIGTGTADGTG 490  
QY 186 KTVUDQKYSVGDTVKTYITYTKNANVYHGTKEVYO-----YV 222  
DB 491 AFTVTVPAGEAGANETLTAVAKNASGTESTPTTFTQPADEATVTAFTTGTGNSVAGY 550  
QY 223 IKDTMPSASVVDLNE-----GSYEVTITDG-SGNIITLTGSEKATGKYNL 267  
DB 551 VAGTADANATVEIRNAGGAVIGTGTADGTGAPTVTPAGEAGANETLTAVAKNASG--- 606  
QY 268 LEENNNFTTITPWAATNTP-----TGNTONGANDDFYKGINTTT 309  
DB 607 -TESTPTTFTQTP-ADPNTPVATPIVETVGTSTTKG-----YEVKGTAEVGTTIEVRDAAGT 660  
QY 310 -----YGVLSKSG-----AKPGSADLPENTNIATINPNT 338  
DB 661 VLGTATTGDKYTVTLDSGTATANTQTLVSVAKNAGSTESQATATTPADVTAPVDNIT 720  
QY 339 SNDDPGQKVT-VRDQGIITIKKIDGSTKASLOGAIF--VLKNATGQF-----LNFNDDT 387  
DB 721 GNSGSGYEITGTADPNNTTIEVRDPS-----GAVIGTGTSDANGDFTVTLPTGTINPGDT 774  
QY 388 NNVT-----EWGTEANATEYTTGADGIIT-----ITGLKEGTYVLVEKKAPLGLYNLLDNSQ 437  
DB 775 LTVIGKDNAGNESQPTVELVPADATVTAFTVGTGNSVAGYQVGTGADPNATIEIRDAD 834  
QY 438 KVLGDGATDTTNS-----NNKGTLPSTGGIGTTI 478  
DB 835 GNVIATGTADGTGSGFAVNLPAGTANANETLTALAKDPAGNTSTPTTFTQPADEVVAPPSV 894  
QY 461 E-----NNKGTLPSTGGIGTTI 478  
DB 895 DKVTGNTTQGYQVGTGAELGTTI 917  
  
RESULT 15  
US-09-071-035-266  
; Sequence 266, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 262:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1638 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-266

```
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 266:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-266

Query Match          7.5%; Score 194; DB 4; Length 1638;
Best Local Similarity 21.8%; Pred. No. 3.3e-06;
Matches 136; Conservative 56; Mismatches 219; Indels 212; Gaps 26;

QY 23 VTPFAFAAETGTTTVDQTKGATYKA---YKVPDAEIDNANVSNDKDGASYLIPQGEA 79
DB 340 VSNIFYAAGTTSVTAPTTGTGNSAGYEVKGTADANATVEIRNAGGT--VIGTG--- 394
QY 80 EYKASTDFNSLFTTTTNGGRTYVTKDTSANEIATWAKSISANT-----TPVSTVT 131
DB 395 ----TADGTGAFTVTVPAGE-----AGANETLTAVAKNASGTEXTPTTQTPADEAT 442
QY 132 ESNNDGTEVINVSQYGVYVSVTVNNGAVIMVTSVTPNATIEKNTDAT-----WGDG-G 185
DB 443 VTAPTITGVGNSTAG-YEVKGTAD-----ANATVEIRNAGGTVIGTGTADGTG 490
QY 186 GKTVDQKTSVGDVTKYTIYKNAVNYHGTKEYQ-----YV 222
DB 491 AFTVTVPAGEAGANETLTAVAKNASGTESTPTTFTQTPADEATVTAPTITGVGNSTAGYE 550
QY 223 IKTMPSASVVDLNE-----GSYEVTTIDG-SGNITTLTGSEKATKYNL 267
DB 551 VKGTADANATVEIRNAGGAVIGTGTADGTGAFTVTTPAGEAGANETLTAVAKNASG----- 606
QY 268 LEENNNFTIIPWAAINTP-----TGNTONGANDDFYKGINITVT----- 309
DB 607 -TESTPTTFTQTP-ADPNVPATPIVETVGSITKG-----YEVKGTAEVGTTEVRDAAGT 660
QY 310 -----YGVKLSG-----AKPGSADLPENTNIATINPT 338
DB 661 VLGTATTTGTDGKVTVLDSGTATANTQLSVAKNASGTESQPATATTPADVTAPTVDNIT 720
QY 339 SNDDPGQKVT-VRDQGITIKKIDGSKASLOGAIF--VLKNATGQF-----LNFNT 387
DB 721 GNSGSGYEITGTADPNNTIEVRDPS-----GAVIGTGTSDANGDFVTPLTGTTPNGDT 774
QY 388 NNW-----EMCTEANATEYTTGADGIIT-----ITGLKEGTYYLVEKKAPLGYNLLDNSQ 437
DB 775 LTVIGKDNAGNESQPTIEVLVPADATVTAPTVGTGNSVAGYQVGTGADPNATIEIRDAD 834
QY 438 KVLGPGADTTNS-----DNLLVNPVT 460
DB 835 GNVIATGTADGTGSAFVNLPAGTANANETLTALAKDPAGNTSTPTTFTQTPADEVWAPPSV 894
QY 461 E-----NNKTELPSTGGIGTTI 478
DB 895 DKVTGNTTQGYQVGTGAELGTTI 917

RESULT 16
US-09-206-942-45
; Sequence 45, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
```

```
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 1095
; TYPE: PRP
; ORGANISM: Haemophilus influenzae
US-09-206-942-45

Query Match          7.5%; Score 192.5; DB 4; Length 1095;
Best Local Similarity 23.8%; Pred. No. 2.4e-06;
Matches 128; Conservative 63; Mismatches 194; Indels 153; Gaps 30;

QY 40 TOKGATYKAYKVFDAEIDNA-----NVSDNKGASLYLIP---OGKEAEYKASTD-FNSLIF 91
DB 511 TNSDSAYRT--IIEGNTNSNGDLNITD-NKNNABIOIGGNISQKEGNLTSSDKINITN 567
QY 92 TTTTNGGRTYVTKDTSANEIATWAKSISANTTTPVSTVTVESNNDGTEVINVSQYGVYV 151
DB 568 QITIKKG---VNKEDSDSST-----ANNANLTTTKELQLTGD---LNIS--GFDKA 611
QY 152 SSTVANGAVIM-----VTSVTPNATIEK-----NTDA-- 179
DB 612 EITAKEGADLLIGSDNNDNNANAKVTFNOVKDSKISADSHNVTLNSKVETSGNNDABS 671
QY 180 TWGDCGGKTVDOKTVSVGDTVKYTIYKNAVNYHGTKEYQYQVVIKDTMPSASVVDLNEGS 239
DB 672 NNGDGTSLTINAKNI---TVNNNITSHKTVNITASENV-----TTKAGTTINATTGS 720
QY 240 YEVTTIDG-----SGNITTLTGSEKATKYNLLENNNFTIIPWAAINTPTGNT 290
DB 721 VEVTAKTGDIKKVSTSGSVTLTATGALAVSNIS-----GNTVTITANKGLTTOAGST 776
QY 291 QNGANDDPFYKGI-----NTIITVYIGVL--KSG-----AKPGSADLPENT 329
DB 777 VSAIN-----GVTASSQSGDISGTSIGNTVKVSAGIDLTTKSGSEIRAKTGEANVTSAT 830
QY 330 NI--ATINPNTSNDPPGQKVTVRDQGITK---KIDG-----STKASLOGA 370
DB 831 GTIGTGISNANV-----VTANTGLTAVEDAADKIDATGGAATLTATSGKLTKASSS-- 882
QY 371 IFVLKNATGQFLNFTNNVEMG--TEANATEYTTGADGIITIG---LKEGTYYLVEK 424
DB 883 ----ITSANNO-VNLSAKDGSIGGNINAAVTLNTTGA--LITVKGSSINANSGLTIVINAK 936
QY 425 KAPLGYNLLDNSQKV-----ILGDGATDTTNSDNLLVNPVTVENNKGTELPSTGGIGTTI 478
DB 937 DAELNGERASGNHTVNVNATNANGSGSVIATTSRNVITGLDITINGLNIISKNGINTVL 994
```

```
RESULT 17
US-09-206-942-43
; Sequence 43, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1101
; TYPE: PRP
; ORGANISM: Haemophilus influenzae
US-09-206-942-43
```



```
Query Match          7.5%; Score 192.5; DB 4; Length 1101;
Best Local Similarity 23.8%; Pred. No. 2.4e-06;
Matches 128; Conservative 63; Mismatches 194; Indels 153; Gaps 30;

QY 40 TQKATYKAYKVFDAEIDNA-----NVSDSKDGASYLIP---QKAEAYKASTD-FNSLIF 91
DB 517 TNSDSAYRT--LIEGNITNSGDLNITD-NKNNABIQGGNISQKEGNLTSSDKINITN 573
QY 92 TTTTNGGRTYVTKDITASANEIATWAKSISANTPVTSTVTSNNDGTEVINVSQGYYYV 151
DB 574 QITIKKG--VNKEDSDST-----ANNANLTIKTKELQLTGD-----LNIS--GFDKA 617
QY 152 SSTVANGAVIM-----VTSVTPNATHEK-----NTDA-- 179
DB 618 EITAKEGADLIIGNSDNNNNNAKVTFNQVKDSKISADSHVTLNSKVTSSNGNDAES 677
QY 180 TWGDDGGKTVDOKTVSGVTYKYNVNHGTEKYVYVVKDTPMSASVVDLNEGS 239
DB 678 NNGDGTSLTINAKNI-----TVNNNITSHKTVNITASENV-----TTKAGTTINATGTS 726
QY 240 YEVTTIDG-----SGNITTLTQGSSEKATGKYNLLENNNFTITIPWAAINTPTGNT 290
DB 727 VEVTAKTGDIKGVESGTSVTLTATGEALAVSNIS-----GNTVITANKGKLTTQAGST 782
QY 291 QNGANDDPFKGI-----NTITVYTGVL--KSG-----AKPGSADLPENT 329
DB 783 VSAIN-----GVTASSQSGDISGTSIGNTVKVSIGAIDLTTKSGSEIKAKTGEANVTSAT 836
QY 330 NI--ATINPTNSDDPGQKVTVRDGGITIK---KIDG-----STKASLOQA 370
DB 837 GTIGGTISGAVN-----VTANTGDLTVEDDAKIDATGGAATLTATSGKLTTKASSS-- 888
QY 371 IFVLKNATQOFLNFDNTNNEVG--TEANATEYTTGADGIITG-----LKEGTYYLVEK 424
DB 889 ---ITSANNQ-VNLSAKDSIGSGINAAVNLNTTGA--LTTVKGSSINANGSLTLVINAK 942
QY 425 KAPLGYNLLDNSQV-----ILGDGATDTNSDNLVNPVTENNKKGTLPSTGGIGTTI 478
DB 943 DAELNGEASGNHTVNVNATNANGSGSVIATTSRNVITGDLITINGLNIISKNGINTVL 1000

RESULT 18
US-09-206-942-28
; Sequence 28, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; EARLIER FILING DATE: 1998-12-08
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1220
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-28

Query Match          7.4%; Score 192; DB 4; Length 1226;
Best Local Similarity 23.2%; Pred. No. 3e-06;
Matches 119; Conservative 61; Mismatches 203; Indels 130; Gaps 23;

QY 55 EIDNANVSDSKDGASYLIPQKAEAYKASTDFNSLFTTTTNGGRTYVTKDITASANEIA 114
DB 666 ETSNGNDAESNNGDS-----TSLTINAKNVTVNNNITSHKTVNITASEN-VT 712
QY 115 TWAKS-----ISANTTPVSTVTSNNDGTEV-----INVSOQYGYVVSSTVNG 158
DB 713 TKAGTTINATIGSVVEVTAKTGDIKGGIESNGSNVNIATASGDTLNVSNITGQNTVVAASG 772
QY 159 AVIMVTSVTPNATHEKNTDATWGGGKTVDOKTVYSGDTPVK-YTITYKNAVNTHGTEK 217
DB 773 AVTTTKGSTINATGNANITTKTGEINGE-VKSASGNVNITASGNTLNVSNITGQNTVVT 831
QY 218 VYQYVVKDTPMSASVVDLNEGSYEVTTIDG-----SGNITTLTQGSSEKATGKYNLL 268

Query Match          7.4%; Score 192; DB 4; Length 1226;
Best Local Similarity 23.2%; Pred. No. 3.1e-06;
Matches 119; Conservative 61; Mismatches 203; Indels 130; Gaps 23;

QY 55 EIDNANVSDSKDGASYLIPQKAEAYKASTDFNSLFTTTTNGGRTYVTKDITASANEIA 114
DB 666 ETSNGNDAESNNGDS-----TSLTINAKNVTVNNNITSHKTVNITASEN-VT 712
QY 115 TWAKS-----ISANTTPVSTVTSNNDGTEV-----INVSOQYGYVVSSTVNG 158
DB 713 TKAGTTINATIGSVVEVTAKTGDIKGGIESNGSNVNIATASGDTLNVSNITGQNTVVAASG 772
QY 159 AVIMVTSVTPNATHEKNTDATWGGGKTVDOKTVYSGDTPVK-YTITYKNAVNTHGTEK 217
DB 773 AVTTTKGSTINATGNANITTKTGEINGE-VKSASGNVNITASGNTLNVSNITGQNTVVT 831
QY 218 VYQYVVKDTPMSASVVDLNEGSYEVTTIDG-----SGNITTLTQGSSEKATGKYNLL 268
```

Db 832 ANSGAI--TTTEGSTINATGCDANITTTQTNINGKVESSGVSFLIATGQTLAVGNIS-- 887  
Qy 269 EENNFTITIPWAAATNPTGNTONGANDFFYKGI-----NTITVTYTG 313  
Db 888 --GDTVIITADKGLITQTSKINGT-----KSVTSSOSGDISGTSIGNTVSVSATGS 939  
Qy 314 L-----KSGAKPGSADLPENTNI--ATINPNTSNDPPGQKTVRDQGITIKKIDGSTKA 365  
Db 940 LTTQAGSKIEAKTGEANVTSATGTIGGTISGNTVN-----VTANTDNLTIK--DGARIK 991  
Qy 366 SLOGAIFVLKNATGQFLNFNDTNNVWG-----TEANATEVTGADGI 408  
Db 992 ATCGAVTL--TATGGTLTTTSSDITSSNGQTLTAKDSSIAGSIANAANVTLLNTTGT--L 1047  
Qy 409 ITITGLK---EGTYLVEKAPLGYNLLDNSQKVLGDGATTTNSDNLNVNPTVENNK 464  
Db 1048 TTVAGSKIEAAGTLVINAKAQL-----DGA---ALGDRTEVNVNANGS 1090  
Qy 465 GTPLPSTGGIGTTFYIIGAILVIGAGIVLVAR 497  
Db 1091 GSVIATK---SSRVNITGLIITN-GLNIISK 1118

RESULT 20  
US-09-134-000C-5999  
; Sequence 5999, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134, 000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055, 778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5999  
; LENGTH: 1747  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5999

Query Match 7.4%; Score 192; DB 4; Length 1747;  
Best Local Similarity 21.7%; Pred. No. 5.1e-06;  
Matches 136; Conservative 59; Mismatches 234; Indels 198; Gaps 26;  
Qy 11 VASLAFGMVSPVTPPIAFAAETGTTIVQDTQKATYKA---YKVFDAEIDNANVSDSKD 67  
Db 439 VAKNASGTESTPTT-FQTPADEATVTA-PTITGVTGNSTAGYEVKGTADANATVEIRNAG 496  
Qy 68 GASYLIPQKAEYKASTDPNSLFTTTTNGRVTYVTKKOTASANEIATWAKSISANTTPV 127  
Db 497 GT--VLGTG-----TADGTGAFVTVPAGEAGANETLTAVAKNASGTESTPTTPTQTPA 547  
Qy 128 STVTESNNDGTEVINVSQYGYVSVSVNNGAVIMVTSVTPNATIEHKNTDAT-----WG 182  
Db 548 DEATVAPTITGVTGNSTAG-YEVKGTAD-----ANATVEIRNAGGTIVGTGTA 595  
Qy 183 DG--GGKTVQDKTSVSGDVKYKTYTKYKNAVNYHGTKEKYQ----- 220  
Db 596 DGTGAFVTVPAGEAGANETLTAVAKNASGTESTPTTFTQTPADEATVAPTITGVTGNST 655  
Qy 221 --VVIKDTMPSASVDLNE-----GSVEVTIDG--SGNITTLTGSEKATG 263  
Db 656 AGEVKTADANATVEIRNAGGAVIGTADGTGAFVTVPAGEAGANETLTAVAKNASG 715  
Qy 264 KYNLLBNNTTITIPWAAATNP-----TGNTONGANDFFYKGINITIVT----- 309  
Db 716 ----TESTPTTQTP-ADNTEPVATPIVETVGTSTKG-----YEVKGTAEVGTTLIEVRD 765  
Qy 310 -----YTGVLKSG-----AKPGSADLPENTNIAT 334

Db 766 AAGTVLGTATTGTDGKYIVTVLLDSTGATANTQTLSSVAKNASGTESQPATATTADVTAPTV 825  
Qy 335 NPNTSNDPPGQKTV--VRDQGITIKKIDGSTKASLOGAIF--VLKNATGQF-----LN 383  
Db 826 DNITGNSGSGYBITGADPNTTIEVRDPS-----GAVIGTGTSDANGDFTVLTPTGTTN 879  
Qy 384 FNDTNNV---BWSGTEANATEYTTGADGLIT-----ITGLKEGTYLVEKAPLGYNLL 433  
Db 880 PGDTLTVIGKNWAGNESQFTEVLVPADATVAPTVTGVTGNSVAGYQVGTGADPNATIEI 939  
Qy 434 DNSQKVLGDGATDTTNS-----DNLIV 456  
Db 940 RDADGNVIATGADGTGSAFVNLPAGTANANETLTALAKDPAGNTSTPTTQTPADEVVA 999  
Qy 457 NPTVE-----NNKGTELPSTGGIGTTI 478  
Db 1000 PPSVDKVTGNTTQGYQVGTGTAELGTTI 1026

RESULT 21  
US-09-206-942-49  
; Sequence 49, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; CURRENT FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 09/167,568  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 49  
; LENGTH: 1073  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-49

Query Match 7.4%; Score 190.5; DB 4; Length 1073;  
Best Local Similarity 21.1%; Pred. No. 3.3e-06;  
Matches 126; Conservative 71; Mismatches 197; Indels 203; Gaps 28;  
Qy 9 LTVASLAFGMVSPVTPPIAFAAE-----TGTIT-----VQDTQKATYKAYKVFDAEI 56  
Db 447 LVGASANTNNLSVKSGAKFKAETNDNLNITGTTTNNGTSIIDVKKGAA-----KLGNITN 502  
Qy 57 D-NANVSDSKDGSYLYIPQKAEYKASTDPNSLFTTTTNGRT-----YVTKKD--- 106  
Db 503 DGNLNTTNNAKQKSVI--NGNITNNKALNI-----TNNNDTEIOIGNISOKENL 555  
Qy 107 TASANEI--ATWAKSISANTTPVSTVTESNNDG-----TEVINVSQYQ 148  
Db 556 TISSDKNITKRIEIKAGT-----DQNSDSGVASNANLTIKTELKLTENLINSFGDK 609  
Qy 149 YVSVSTVNGAVI-----MVTSTPNATIEHK-----NTDA 179  
Db 610 ABEVAKENNLIIIGNNGDNANAKTVTFNNVSKSISANGHNVTILNSKVETSDGNSNTEG 669  
Qy 180 TWGDCGGKTVQDKTSVSGDVKYKTYTKYKNAVNYHGTKEKYQVVIKDTMPSASV----- 232  
Db 670 NSDNNAGLTIDAKV-----TVNNDITSHKTVNITASERI-----DTKADTTINATGNV 719  
Qy 233 -----VDLNEGSYEVTIDG-----SGNIT-----TLTGSEKATGKY 265  
Db 720 KLTAVTSDIQGKISNGSDVNITSTGSKINGKESGVSILTATEKLTGVN----- 772  
Qy 266 NLENNNTTITIPWAAATNP-----YTGVLKSG-----AKPGSADLPENTNIAT 334

Db 773 ---VSGNTVTVTANRGAITTLTLAGSTINGTN-----GVTTSSQSGEIGGEVTKTVSVTA 823  
Qy 311 TG---VLKSGAKPGSADLPENTNIATINPNTSNDPGQKVVTRDQGITIKKIDGSTKASL 367  
Db 824 TAGSLTVKGGAK---INATEGTATLTASSGKLTTEASSNITSAGKQVDLSAQDGS-----875  
Qy 368 QGAIFVLKKNATQGFNLFNNDTNVWKGTEANATEYTTGADGIITIG-----LKEGTYYLVE 423  
Db 876 -----IAGQI-----SAANVTLTNTGT--LTTVEGSSINANEGLTVINA 912  
Qy 424 KKAPLGYNLLDNSQV-----ILDGDGATDTTNSDNLNVPNTVENNKGTELPSTGGIGT 476  
Db 913 NDAKLDGKASGRNTEVNATNAGSGSVTAKTSSSVNITGDLNTINGLNIISENGRT 969  
RESULT 22  
US-09-206-942-47  
; Sequence 47, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loomore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:Jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; CURRENT FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 09/167,568  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 47  
; LENGTH: 1079  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-47  
Query Match 7.4%; Score 190.5; DB 4; Length 1079;  
Best Local Similarity 21.1%; Pred. No. 3.3e-06;  
Matches 126; Conservative 71; Mismatches 197; Indels 203; Gaps 28;  
Qy 9 LLVASLAFGMVSPVTPPIAFAB-----TGTIT-----VDQKQATYKAYKVFDAEI 56  
Db 453 LVGASANINNLNVKSGAKFAETNDNLNITGTFTNNGTSIIDVKKGA-----KLGNITN 508  
Qy 57 D-NANVSDNSKDGASVLIPOGKEAEKASTDFNSLFTTTTNGRT-----VYTKD--- 106  
Db 509 DGNLNTTNNAKQKQSVI-NGNITNNKALNI-----TNNGNDTEIQIGGNISQKGNL 561  
Qy 107 TASANEI-ATWAKSISANTTPVSTVTESNNDG-----DOGNSDGSVANANLTIKTELKLTENLINSQPK 515  
Db 562 TISSDKINIKRIEIKAGT-----DOGNSDGSVANANLTIKTELKLTENLINSQPK 515  
Qy 149 YVVSSTVNNGAVI-----MVTSTVTPNATHEK-----NTDA 179  
Db 616 ASIVAKENNNLIIGNNGDNANAKVTFNNVVKDSKISANGHNVLNLSKSVETSDGNSNTEG 675  
Qy 180 TWGDDGGKTVDOKTSYVSGVTYKTYTYKAVNVHGTKEYVQYVVKDTPMSASV-----232  
Db 676 NSDNNAGLTIDAKV-----TVNNDITSHKTVNITASERI-----DTKADTITNATNGV 725  
Qy 233 -----VDLMEGSYEVTITDG-----SGNIT-----TLTQSEKATGKY 265  
Db 726 KLTAVTSDIQGGKNSGDVNITTSNGSINGKIESKSGSVTLTATEKTLTVGN-----778  
Qy 266 NLLSENNFTIIPHAATVTPGTONTONGANDOFFYKGI-----NITVTY 310  
Db 779 ---VSGNTVTVTANRGAITTLTLAGSTINGTN-----GVTTSSQSGEIGGEVTKTVSVTA 829  
Qy 311 TG---VLKSGAKPGSADLPENTNIATINPNTSNDPGQKVVTRDQGITIKKIDGSTKASL 367  
Db 830 TAGSLTVKGGAK---INATEGTATLTASSGKLTTEASSNITSAGKQVDLSAQDGS-----881

Qy 368 QGAIFVLKKNATQGFNLFNNDTNVWKGTEANATEYTTGADGIITIG-----LKEGTYYLVE 423  
Db 882 -----IAGQI-----SAANVTLTNTGT--LTTVEGSSINANEGLTVINA 918  
Qy 424 KKAPLGYNLLDNSQV-----ILDGDGATDTTNSDNLNVPNTVENNKGTELPSTGGIGT 476  
Db 919 NDAKLDGKASGRNTEVNATNAGSGSVTAKTSSSVNITGDLNTINGLNIISENGRT 975  
RESULT 23  
US-09-071-035-264  
; Sequence 264, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION NUMBER: US/09/071.035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 264:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 669 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-264  
Query Match 7.4%; Score 190; DB 4; Length 669;  
Best Local Similarity 22.5%; Pred. No. 1.8e-06;  
Matches 122; Conservative 60; Mismatches 242; Indels 118; Gaps 22;  
Qy 11 VASLAFGMVSPVTPPIAFABETGTTVDQKQATYKA---YKVFDAIDNANVSDSKD 67  
Db 13 VAKASGTEXTPTT-FQTPADENVTA-PTTGTGTGNTAGYEVKGTADANATVEIRNAG 70  
Qy 68 GASVLIPOGKEAEKASTDFNSLFTTTTNGGRTYVTKDITASANEIATWAKSISANTTPV 127  
Db 71 GT--VIGTG-----TADTGAFVTVPAGEAGANETLTAVAKNASGTESTPTTFTQTPA 121  
Qy 128 STVTESNNDGTEVINVSQYGYVYVSSTVNNGAVIMVTSVTNATHEKNTATWGDGSGK 187  
Db 122 DEATVAPTITGTGNTAG-YEVKGTADANATVEIRN-AGGAVIGTGTADGT-----GAF 175  
Qy 188 TVDQKTSYVSGVTYKTYTYKAVNVHGTKEYVQ-----YVI 223  
Db 176 TVTIPAGEAGANETLTAVAKNASGTESTPTTFTQTPADNTPVATPVTGTTGTTKGYEV 235

QY	224	KOTMPSASVDLNEGSYEV-----TITDGSNITTLTQGSSEKATGKNLLENNNFITI	278
Db	236	KGTAEVGTTIEVRDAAGTVLGTATGTGKYTV-TLDSGTATANQTLSSVAKNASGTESQ	294
QY	279	PWAATNTP-----TCNTQNGANDFFVKGINITITVTVT-----GVLKSGAK	319
Db	295	PATAT-TPADVTAFTVDNITGSSG-----YEITCTADPNTTIEVRDPGAVIGTETS	347
QY	320	PGSADLPENTNIATNP-----NTSND-DPGOKVTVRDCQITIKKIDGSTKASLQ	368
Db	348	DANGDFTVTLPTGTNPGLTIVIGKDNAGNESQTEVLVPADATVTAFTVTGVTGNSVA	407
QY	369	GAIFVLKNATCGQLFNFDNTNNVEWGTENATEVTTGADGIIITITGLKEGTYVLVEKKAPL	428
Db	408	G-----YQVTGT-----ADPNATIEIRDADGNVIATGTADGTGSEA-VNLPA	448
QY	429	CYNLLDNSQKVLIDGADTDT-----TNSDNLLVNPTVE-----NNKGTELPSTGGIGT	476
Db	449	GTANANETLTAALAKPAAGNTSTPTTFOTPADEVVAPESVDKVTGNTTQGYQVGTABLGT	508
QY	477	TI 478	
Db	509	TI 510	

RESULT 24  
 US-09-386-962C-4  
 ; Sequence 4, Application US/09386962C  
 ; Patent No. 6635473  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FOSTER, Timothy  
 ; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM CONGULASE-NEGATIVE STAPHYLOCOCCUS EPIDERMIDIS  
 ; FILE REFERENCE: F06335US2/GAS  
 ; CURRENT APPLICATION NUMBER: US/09/386,962C  
 ; CURRENT FILING DATE: 1999-08-31  
 ; PRIOR APPLICATION NUMBER: 60/098,443  
 ; PRIOR FILING DATE: 1998-08-31  
 ; PRIOR APPLICATION NUMBER: 60/117,119  
 ; PRIOR FILING DATE: 1999-01-25  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: Patentin version 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1742  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-386-962C-4

[illegible]

```

701  -----KNGVQDD--DEKGLAGVYVTLKDSNNRELQRVTTDOSGHYQFDNLQNG 740
318  AKPGSADLPENTNIATINPNTSND-----DPGQKVTVRDGQITKKIDGST----- 363
747  TTYTVEFALPDNYTSPAN-NSTNDAIDSDGERDGRKVVVAKG--TINNADNMTVDTGFY 803
364  -----KASLQGAIFVLKNATGQFLNPNFTNNVWEGTEANAT 399
804  LTPKYNVDYVWEDTNKGIDQDDNEKGISGVKVTLLKNKNGDTI-----GTT----- 849
400  EYTTGADGIITITGLKSGTYVLYVEKKAPLGY-----NLLDNSQ 437
850  --TTDSNGKYEFTELENGD-YTISFETPEGYTPTKQNSGSDGKDSNGTKTKTIVTKDADN 906
438  KVI-----LGDGATDTNSDNL-----VNPTVENNKGTPELSTGGIGTT 477
907  KTIDSGFKPYNLGDYWEDTNKGIDQDDSEKGISGVKVTLLKDKNG-----NAIGTT 959

RESULT 25
US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
; GENERAL INFORMATION:
;   APPLICANT: Smit, John
;   APPLICANT: Bingle, Wade H
;   TITLE OF INVENTION: Bacterial surface protein expression
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Shlesinger, Arkwright & Garvey
;   STREET: 3000 South Eads Street
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: USA
;   ZIP: 22202
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/194,290
;   FILING DATE: 09-FEB-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Garvey, George A
;     REGISTRATION NUMBER: 17737
;     REFERENCE/DOCKET NUMBER: 5946-1
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 703-684-5600
;   TELEFAX: 703-836-5288
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1026 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-194-290-7

```

	Query Match	7.2%	Score 185.5;	DB 1;	Length 1026;
	Best Local Similarity	22.1%;	Pred. No. 7.6e-06;		
	Matches 131;	Conservative	66;	Mismatches 240;	Indels 157; Gaps 26;
Qy	6 IQSLLVASLAFGMVSPVTPIAFAEFGTTTIVDQTQKGYAKYKVPDAIDNANVSDSN	65	:	:	:
	: : : : : :	:	:	:	:
Dd	299 VQAAAVATLPGTVTISGIE--TNVTISGAALITLTSSGVT-----GLTALNTN	344	:	:	:
	: : : : : :	:	:	:	:
Qy	66 KDGASYLIPOGKBAEYKASTDFNSLFTTTNGG-----RTYYTK-KDTASANEIATWAK	118	:	:	:
	: : : : : :	:	:	:	:
Dd	345 TSGAAQTIVTAGAGQLTATTAAQAANNVAVDGGANVTVASTGVTSGTTTVGANSAAASGV	404	:	:	:
	: : : : : :	:	:	:	:
Qy	119 SIS-ANTTPVST-----VTESNDGTGEVINVSQYG	148	:	:	:
	: : : : : :	:	:	:	:

Db 405 SVSVANSSTTTTGAIAVGTGTAATVVAQTAGNAVNTTLTQADVTVTGNSSTTAVTQTAA 464  
QY 149 YVVSSTVN---NGAVIMVTSVTPNATIHENKNTDATTGDCGGKTVQDKTVSGDVTVKYIT 205  
Db 465 ATAGATVAGRVNGAVTITDSAAASATTAGKATVTLGSGCAATIDSSALT-----514  
QY 206 YKNAVNYHTEK---YVQVVKDTPSASVVDLNEGSYEVT---ITDGG---NITTLTQ 256  
Db 515 ---TVNLSGTGSLGIRGALTIAT-PTANTLTUNVGLTTGAITDSEAAADGGFTTINI 570  
QY 257 GSEKATGKYNLLENNFTITIPWAAATNTPTGNTQNGANDDFYKGINITVT-----309  
Db 571 AGSTASSTIASLVAADATLINISGDARVITISHTA-----AALTGITVNSVATL 621  
QY 310 ----YTGVLKSGAKPGSADLPENTNIA-----TINPNTSNDPQGVKTVRDG-QIT 356  
Db 622 GAELATGLVFTGAGADSIILLGATTKAIVMGAGDDTVTVTSATLGGGSGVNGDGDVLV 681  
QY 357 KIDGSTKASLOGAIFVLKXNATQF-----LNENDTNVNGTEANATEYTT 403  
Db 682 ANVNGSS-----PSADPAFGFETLVRVAGAAAGSHNANGFTALQIGATAGATTFTN 733  
QY 404 GADGI-ITITGLKEGYIYVEKAPLG---YNL-LDNSQKVLGD-----GAT 446  
Db 734 VAVNGLTVLAAPTGTITVTLANA-TGTSDFVNLTLSSAALAGTVALAGVETVNIAT 792  
QY 447 DTNNS---DNLLVNP-----VENNKGTELPSTGGIGTTFIYIGAILVIGAG 491  
Db 793 DTNTTAHVDTLTLQATSAKSIYVVGAGLNLNTGNTAVTSF---DASAVTGTG 843

RESULT 26  
US-09-252-991A-30227  
; Sequence 30227, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30227  
; LENGTH: 2736  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30227

Query Match 7.1%; Score 183.5; DB 4; Length 2736;  
Best Local Similarity 20.7%; Pred. No. 4.5e-05;  
Matches 112; Conservative 73; Mismatches 222; Indels 133; Gaps 21;

QY 21 SPVTPIAFAAETGTI---TVQDTQKATYKAYKVFDA-----EIDN-ANVSDSNKDG 68  
Db 1066 TPSTPLA---DGTVNATATDPAGNTGGSTTVDAIAPATPTVNLSSGSLSGTAEPG 1121  
QY 69 ASYLIPQKAEYKASTDFNSLFTTTNGGR---TYVTKKDTASAEIATWAKSISANTTP 126  
Db 1122 STVILTGD-----NGNPIAEVATDGGSNWTYTPSPIANGTVVNVVAAADAGNSSP 1172  
QY 127 VSTVTT-ESNNDGTEVINVSQYGYVVSSTVNGAVIMVT-----SVTPNATHEKNT 177  
Db 1173 PATVTVDSSAPPAPVINS-NGWISGTAEGATVTLTDAGNPIGQVATDAGSGNWSFT 1230  
QY 178 DATWBGCGKTVQDKTVSGDVTVKYITITVKNANVYHTEKTVQYVVKDTPMPSASVVDLNE 237  
Db 1231 PGTPPANGTIVATATDPTGNTGPQAATTVDAV-----APPAPVIDPSN 1274

QY 238 GS-----YEVITDGGSGNITTLTQSEKATCKYNLLENNFTITIPHAATNPTTG 288  
Db 1275 GTTISGTAEAGAKVILTDGNG--PIGETTADGSGNWTFTPATPLANGTVVNAVAQDPAG 1332  
QY 289 NT--QNGANDDFYKGINITITVTVGLKSGAKPGS-----322  
Db 1333 NTGPGSITVDVAVAPNTPVNPSNGNLLNGTAEPGSTVTLTDGNGNPIGQTTADSGNWS 1392  
QY 323 ---ADLPENTNI-ATINPNTSNDPQGVKTVRDGQITIKKIDGSTKASLOGAIFVLKNA 377  
Db 1393 FTPGSQLPNGTVVNTASDAAGNTSAPATTVTDSSLSPIQVDPSPNGSVISGTA-----1446  
QY 378 TGQFLNFNDTNVNGTEANAT---EYTTGADGIITIT---GLKEGTYVLYEKKAPLGN 431  
Db 1447 -----DAGNTIITIDGNGNPIGQVATDAGSGNWSFTPGIPLDGTVVNVVARSF---S 1495  
QY 432 LLDNSOKVILGDGATDTTNSDNLLNVPVENNKGTELPSTGGIGTTFIYIGAILVIGAG 491  
Db 1496 NVDSAPAVITVDGVAPAA-----PVIDPSNGTEISGTAEGATV-----ILTDGGG 1541

RESULT 27  
US-09-268-347-36  
; Sequence 36, Application US/09268347  
; Patent No. 6335182  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
; FILE REFERENCE: 1038-860  
; CURRENT APPLICATION NUMBER: US/09/268,347  
; CURRENT FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 36  
; LENGTH: 2411  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-268-347-36

Query Match 7.1%; Score 183; DB 4; Length 2411;  
Best Local Similarity 22.1%; Pred. No. 4.1e-05;  
Matches 124; Conservative 71; Mismatches 194; Indels 172; Gaps 30;

QY 32 TGTITVQDTQKATYKAYKVFDA--EIDNANVSDSNKDGASYLIPQKAEYKASTDFNS 89  
Db 190 SSSFTPNDEKTRAATVKDVLNAGWNKGAKTAGGNSVD-LVSAYNNVEF-ITGDKNT 247  
QY 90 L---FTTTTNGGRVTV--TKKDTA---SANEIATWAKSISANTTPVSTVTSNNNDGTEVI 141  
Db 248 LDVVLTAKEGKTEVKTPTKTSVIKEXDGLFTGKENNDTNKVTSTNTATDNTDEGNLV 307  
QY 142 N-----VSQVGYVVSSTVN--NGAVIMVTSVTPNATIHENKNTDATWDCGGKTVDQK 192  
Db 308 TAKAVIDAVNKAQWRVKTITANGQGDPAVTSAGT-VNTPES-----GDGTTASVTKD 359  
QY 193 TYSVGDVTKYITTYKNAVNYHTEKTVQYVVKIDTPMPSASVVDL-----NEGSEYV 242  
Db 360 TNGNGITVKYDAKVGDLGKFDSDKKIVADTTALTITVGKVAEIAEKDDKKLVNAGDLVT 419  
QY 243 TI-----TDGSGNITTLTQSGE-----KATGKYNLLENNFTITIPWA 281  
Db 420 ALGNLSWKAKEADTTDGALEGISKDQEVKAGETVTFPKAGKNLKVQDGANFTYSLQDA 479  
QY 282 ATNPTT---GNTONGAND-----DFFYKGINITITVTVGLKSGAKPGS 322  
Db 480 LTGLTSITLGGTTNGNDKAVTKVNDKGLTITPAGNGGTTGNTISVT-----KDGIKAGN 534  
QY 323 ADLPENTNIAT-----INPNTSNDP-----GQ 345  
Db 535 KAI---TNVASGLRAYDDANFDVLNNSATLNRHVEDAYKGLLNLENNKANKQPLVTDST 591  
QY 346 KTVTRDQGITTKKIDGSTKASLOGAIFVLKXNATQFNLNNDNVEWTEANATEYTTGA 405

Db 592 AATVGD-----LRKL-----GWWVSTKNGTKE-----ESNQVKQADEV-----LFTGA 629  
QY 406 DGIITITGLKEG-----TYLVKKAAPLGYNLLDNSOKVILGDGAT-----DITNDSNLL 455  
Db 630 -GAATVTSKSENGKHTITVSVAETKADSG-----LEKGDGTIKLVQNDQNDVL 678  
QY 456 VNPTVENKGTGELPSTGGIGT 476  
Db 679 ---TVGNN-GTAV-TKGGFET 694

RESULT 28  
US-09-328-352-5503  
; Sequence 503, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5503  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5503

Query Match 7.1%; Score 183; DB 4; Length 3892;  
Best Local Similarity 23.8%; Pred. No. 8.2e-05;  
Matches 149; Conservative 69; Mismatches 218; Indels 190; Gaps 35;

QY 12 ASLAFGMVSPVPIAPFAAS-TGFTITVQDTQKATYKAYKVDAEi---DNANVSPNKD 67  
Db 240 ANAVTVINGVYTNATVKAAGTWIVSPGSLVADADKTIIDAKVFTFDAGNSSVND 299  
QY 68 GASVLI-----POGKEAEYKASTDFNSLFTTTTNGRTYVTKDTASANEIA-----TWAKS 119  
Db 300 TQTVTLDTAPNAPVIDPVNGTD--PITGTAEPGSTVTVTPYDGSCTKVVAGPDGTW---354  
QY 120 ISANTTPSVTTSNNDGTEVINVSQYGYVSVSTVNGGAVIMVTSVTPNATHEKNT-D 178  
Db 355 ----TVP-----NPLNDGDGEVTAAT-----DPAGNTSGPATVADVAVPTVALDVLTD 402  
QY 179 ATWGDGGCKTVDOKTVSGVDTVKYTIYKNAVNYHGTEKYVQVVKD--TWPSASVVDLNE 237  
Db 403 STPALTG--TVNDETAIVVNVVD-GVDYPAVNGDGT-----WTLADNTLPT-----LAD 449  
QY 238 GSYEVTI--TDGSGNI-----TTLTQSEKATGKYNLLENNNFTIT 277  
Db 450 GPHTITVTATDAAGNVGTDGVTVDTPAAPTAGVTFTIDSVTADNVINASEAGNVIT 509  
QY 278 -----IPWAATNTP-----TGNTONGANDDFPKGINITITVTVGLKSGAK-----319  
Db 510 GVLKNIPADATNATVTVINGVYTNATVDK--TAGTWTVSPGSLVADADKTIIDAKVTF 567  
QY 320 ---PGSADLPENTNIAFI-----NPNSTNDPDGQKTV-----349  
Db 568 TDAAGNSSTVNDTQIYTLDTAAPAAPVIDPVNGTDPITGTAEPGSTVTVTPYNGDPTATVV 627  
QY 350 -----RDG-QITIKKIDGSTKASLQGAIFVLKNATG---QPLNFNDTNVVE 391  
Db 628 AGPDGWSVNPGLNDGVEAEIATDPAGNPSLPGTATV--DAVGNTDGVNFT--VDSVT 684  
QY 392 WGTAEANATETTGADGIIITGLKEG-----TY-YLVEKKAPL-----428  
Db 685 ADNVINASE-----ASGNVTIVGLKNVPADAAANTVTVWINGQTYATVDSAGTATVSV 740  
QY 429 -GYNLLDNSOKVI-----LGDGAT-----DITNSNLLNVNPTVENKGTGELP 469  
Db 741 PGSDLTADADKTIIDAKVFTDAAGNSSVNDTQTYITDITAPAPVINPV-----NGTD-P 795

QY 470 STG----GIGTTIFYIIGAILVIGAG 491  
Db 796 IIGTAEPGSTVTVTPYDGSITTTVWAG 821

RESULT 29  
US-09-200-650E-5  
; Sequence 5, Application US/09200650E  
; Patent No. 6680195  
; GENERAL INFORMATION:  
; APPLICANT: Patti, Joseph M.  
; APPLICANT: Foster, Timothy J.  
; APPLICANT: Hook, Magnus A.O.  
; APPLICANT: Bidhinn, Deirdre Ni  
; APPLICANT: Perkins, Samuel L.  
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus  
; FILE REFERENCE: P06283US2/BAS  
; CURRENT APPLICATION NUMBER: US/09/200,650E  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,815  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: 60/098,427  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1315  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-200-650E-5

Query Match 7.0%; Score 181.5; DB 4; Length 1315;  
Best Local Similarity 22.0%; Pred. No. 2.2e-05;  
Matches 116; Conservative 72; Mismatches 191; Indels 149; Gaps 29;

QY 34 TTVQDTQKATYKAYKV--FDAETDNNVSDNSDKGAS---YLIPQKKEAB--YKAST- 85  
Db 427 TTVNPSSENSLTNAKLKVQAHSSYPN-NIGQINKDVTDIKIYQVPGKYTLNKGVDVNTK 485  
QY 86 -----DFNSLFTTTTNGRTVV-----TKDTSANESIAIWA--- 117  
Db 486 ELTDTVNTYQKLTITGDNNSAVIDFGNADSAIVVMVNTKFKQYTNSESTLVQMATLSSTG 545  
QY 118 -KSISANTTPSVTTSNNDGTEVINVSQYGYVSVSTVNGGAVIM---VTSVTPNATI 172  
Db 546 NKSVSSTGNALGFTNNQSGAGQEVYKI---GNYVMEDTNKNGVQELGKGVGNVT--VTV 600  
QY 173 HERNTDATWGDGGKTVDOKTVSV-----GDTVKTIITYKN-----208  
Db 601 FDNNTNTKVGE--AVTKEDGSYLIPNLPGD---YRVEFSNLPKGYEVPSPKQGNNEELD 655  
QY 209 -----AVNYHGTEK-----VYQ-----YVVKDTMPSASVVDLNE---GSYEVTID 246  
Db 656 SNGLSASVTVNGKDNLSADLGIYKPKYNLGDVWEDTNNK-GIQODEKIGISGVTVTLKD 714  
QY 247 GSGNITTLTQSEKATGKYNLLE-ENNNTFITIPWAATNTPGNTQNGANDDFPKGINI 305  
Db 715 ENGNV--LKTVTTDADGKYKFTDLONGYKVEFTTPEGVTPITVT--SGSDIEKDSNGLTT 771  
QY 306 IIVTVTVGLKSKAKPGSADLPENTNIAINPNTSNDPDGQKV---TVRDGQITIKKIDGS 362  
Db 772 -----TGVI-NGA-----DNMTLDSGFYKTPKYNLGNVWEDTNKDGK-----QDS 811  
QY 363 TKASLQGAIFVLKNATGQFLNFDNTNNVWEGTEANATEYTTGADGIIITGLKEGTYLV 422  
Db 812 TEKGISGVTVTLANENGVEVLQTTKDK-----DQYQFTGLENGT-YKV 854  
QY 423 EKKAPLGYNLLDNSOKVILGDGATDTNNSDNLNVNPTVENKGTGELPS 470  
Db 855 EFETPSGYT-----PTQVSGTDEGIDSGNSTTGTGVIKDKDNDTIDS 896

RESULT 30

US-08-621-944A-4  
; Sequence 4, Application US/08621944A  
; Patent No. 6440425  
; GENERAL INFORMATION:  
; APPLICANT: SASAKI, Ken  
; APPLICANT: HARKNESS, Robin E.  
; APPLICANT: LOOMORE, Sheena M.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER  
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/621,944A  
; FILING DATE: 26-MAR-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/478,370  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-587  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1833 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-621-944A-4

Query Match 7.0%; Score 181.5; DB 4; Length 1833;  
Best Local Similarity 23.7%; Pred.No.3.6e-05;  
Matches 136; Conservative 59; Mismatches 205; Indels 173; Gaps 31;  
QY 17 GMAVSPVTPFAFAAETGTTITVDQTKGATYKAYKVPDAEIDNANVSDSNKDGASYLLPQG 76  
DB 351 GISVTP-TEISVDAKSGNV-----APTYNI-----GVKTELNSDGT----- 388  
QY 77 KEAEYKASTDFNSLFT-----TTNGRGTYYTK-KDTASANEIATWAKSISA 122  
DB 389 DKFSVKSGTNSLSVTAELASLYLNEVNRVTADSALQSFTVKEEDDDANAI-TVAKDITK 447  
QY 123 NITPVTV-----TESNNDGTEVINVSQ-YGYVYVSVTVNGAV-----IMV- 163  
DB 448 NAGAVSILKLGKNGLTVAATKKDGTGVLGSLQSDGLTIKSTLNNDGLTVKDTNEQIQVG 507  
QY 164 -----TSVTPNATIEKNTDATWGGGKT-----VDQKTYSGVDTVKYTI 204  
DB 508 ANGKFTNVNGSPGFGIANTARIITDKIGFAGSDGAVDTNKPFLDQDKLVGN-VKITN 566  
QY 205 TYQNANVYHGTKEYQYVVKDTPMSASVVDLNEGSYEV--TITD-----GSGNITTLTQGS 258  
DB 567 TGINA---GGRAITG--LSPTLP--SIADQSSRNIELGNTIQDKKSNAAASINDILN-- 616  
QY 259 EKATGKNVLENNNFTIIPW-----AATNPTGNTONGANDFFYKGINITVT 309  
DB 617 ---TG-FNLKNNNPIDFVSTVDIVDFANGNATTATVTHDTANKTSKVYDVNVDDTTIH 672

QY 310 YTGIV---LKSGAKPGSAD-LPENTNIATINPNTSNDPQGVTVRD----- 351  
DB 673 LTGTDNKKLGKVTTKLNKTSANGNTAT-NFNVNSDEEDALVNAKDIAENLNTLAKEIHT 731  
QY 352 -----GOITIKIDGSKASIQGAIFV-LKNATGQF--LNFNDTNNVWEGTEANAT 399  
DB 732 TKGTADTALQFTTVKVDENNADANAITVGQKANNQVNTLTILKGENGLNIKTDKNGT 791  
QY 400 EYTTGADGIIITIGLKEGTYVYLVEKKAPLGYNLLDNSQKVLGDCGATDTTNSDNLVNPT 459  
DB 792 V-----TFGINTTSLKAG-----KSTLNDGGL-----S 815  
QY 460 VENNKGTELPSTGGIGCTTIFYIIGAILVIGAGI 492  
DB 816 IKNPTGSEIQVGDGVK-FAKVNNNGVVGAGI 847

Search completed: July 22, 2004, 10:29:49  
Job time : 24 secs



